NAS9-14016
T-1039
MA-129TA
E76-10470
NASA CR-
CR-147856

FINAL REPORT

(E76-10470) [PHOTOINTERPRETATION OF SKYLAB IMAGERY] Final Report, 1 Jun. 1975 - 31 May 1976 (Purdue Univ.) 278 p HC $9.25 CSCL 05B
N76-32607
Unclas
G3/43 00470

NASA Contract NAS9-14016

June 1, 1975 - May 31, 1976

D. A. Landgrebe, Purdue University
Principal Investigator

A. E. Potter, NASA/JSC
Technical Monitor

Submitted by

The Laboratory for Applications of Remote Sensing
Purdue University, West Lafayette, Indiana

1976

Table of Contents

RESEARCH TASKS

Preface

This report provides a summary of results for the second year's effort under contract NAS9-14016. The contract called for work on a wide variety of separate and distinct but related tasks. Below is a list of the tasks as contained in the original Work Statement.

As a result of this contract a large volume of results has been generated. Technical and research reports previously submitted or presently being published are listed in Table I.

Because of the diversity present in the task list for the contract, each major subdivision of this report has been written to be relatively self-contained. We hope this will facilitate use of the report by readers with different interests.

The various tasks have been managed by various Purdue staff members during the year and a NASA-appointed task monitor was associated with each. It is appropriate that the contributions of these people be recognized.

Task	Purdue Task Manager	NASA Task Monitor
Exhibit D		
2.1 Layered Classifier Adapted to Multitemporal Data Sets	P.H. Swain	Kenneth Baker
2.2 Development of Signature Extension Strata from Clustering Techniques	M.E. Bauer	J.G. Garcia
2.3 Field Measurements Research for Remote Sensing of Wheat	M.E. Bauer	Michael McEwen
2.4 Scanner System Parameter Selection	L.F. Silva	Kenneth Demel
2.5 Transfer of Computer Image Analysis Techniques	J.C. Lindenlaub	J.D. Sargent
2.6 Research in Remote Sensing Technology	P.A. Anuta	A.E. Potter
2.7 Forestry Applications of Computer Aided Analysis Techniques	R.P. Mroczynski	Linwood Smelser, USI
2.8 Analysis of Texas Coastal Zone Environments	R.A. Weismiller	G.E. McKain
Exhibit B		
EROS Data Center Tasks	T.L. Phillips	Donald Orr

The efforts of Dr. A.E. Potter, the contract Technical Monitor are especially to be noted and greatfully acknowledged.

TABLE I. Technical and Research Reports

052576 Lindenlaub and Lube. Matrix of Educational and Training Materials in Remote Sensing.

052076 Mroczyński, Goodrick, Berkebile and Scholz. Analysis of Aircraft MSS Data for Timber Evaluation.

051576 Bauer and Davis. Stratification of LANDSAT Data by Clustering.

051676 Mroczynski. Application of Satellite Collected and Computer Analyzed Data to the Management of the Central Hardwood Forest.

030176 Bauer. Technological Basis and Applications of Remote Sensing of the Earth's Resources.

012376 Berkebile, Russell and Lube. A Forestry Application Simulation of Man-Machine Techniques for Analyzing Remotely-Sensed Data.

012176 Bizzell, Hall, Feiveson, Bauer, Davis, Malila and Rice. Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS) Project.

100675 Anuta. Computer-Aided Analysis Techniques for Remote Sensing Data Interpretation.

100175 Peterson, Goodrick and Melhorn. Delineation of the Boundaries of a Buried Pre-Glacial Valley with LANDSAT Data.

072475 Fleming, Berkebile and Hoffer. Computer-Aided Analysis of LANDSAT-1 MSS Data: A Comparison of Three Approaches, Including a "Modified Clustering" Approach.

072175 Bauer, Cary, Davis, and Swain. Crop Identification Technology Assessment for Remote Sensing (CITARS).

063675 Peterson. Quantitative Inventorying of Soil and Land Use Differences by Remote Sensing.

062775 Phillips, Grams, Lindenlaub, Schwingendorf, Swain and Simmons. Remote Terminal System Evaluation.

062375 Kettig and Landgrebe. Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects.

061875 Parsons and Jurica. Correction of Earth Resources Technology Satellite Multispectral Scanner Data for the Effect of the Atmosphere.

061275 Swain, Wu, Landgrebe, and Hauska. Layered Classification Techniques for Remote Sensing Applications.

OTHER REPORTS

2.1 Layered Classifier Adapted to Multitemporal Data Sets

INTRODUCTION

Rationale

Most contemporary systems which use classification in analyzing multispectral remote sensing data employ classifiers which consist of a single stage of decision logic. As depicted schematically in Figure 2.1-1, to classify each point in the data, these classifiers use a single set of features (multispectral measurements) to compute a likelihood or discriminant function for each class and categorize the point according to the class with the largest discriminant function value. For many practical applications, and particularly those utilizing multitemporal data, such a simple decision procedure is often found seriously lacking in the flexibility needed to most efficiently and effectively incorporate the multitemporal aspects of the data.

As a simple illustration, consider a situation in which data from two passes of a satellite are available over an area to be classified. The conventional single-stage classifier treats the "stacked" combination of the two four-dimensional vectors (one vector from each pass) as a single eight-dimensional vector. If for any reason the data in one or more of the channels is "invalid," the classifier will probably not be able to assign the data to the correct class. For example, if on either date, the ground at a given point happens to be cloud-covered, the classifier would not be able to correctly identify the ground cover. If pass #1 has C_1 percent cloud cover and pass #2 has C_2 percent, as much as $C_1 + C_2$ percent of the multitemporal data may not be classifiable.

On the other hand, a layered classifier, designed as shown in Figure 2.1-2 to make a sequence of decisions, alleviates the problem significantly. Using this classifier the only points which cannot be classified at all are those obscured by clouds on both passes.

This illustration is but one representative of a wide range of classification problems which are most effectively solved by means of layered decision logic. Several other examples are noted below:

*Multitemporal data

-Training drawn from multiple areas

-Change detection

*Use of mixed data types

-Spectral/topographic data

-Spectral/textural data

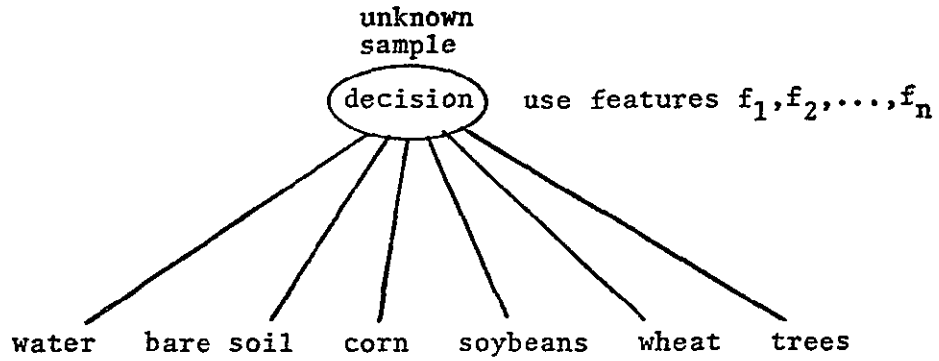


Figure 2.1-1 Common single-stage classifier.

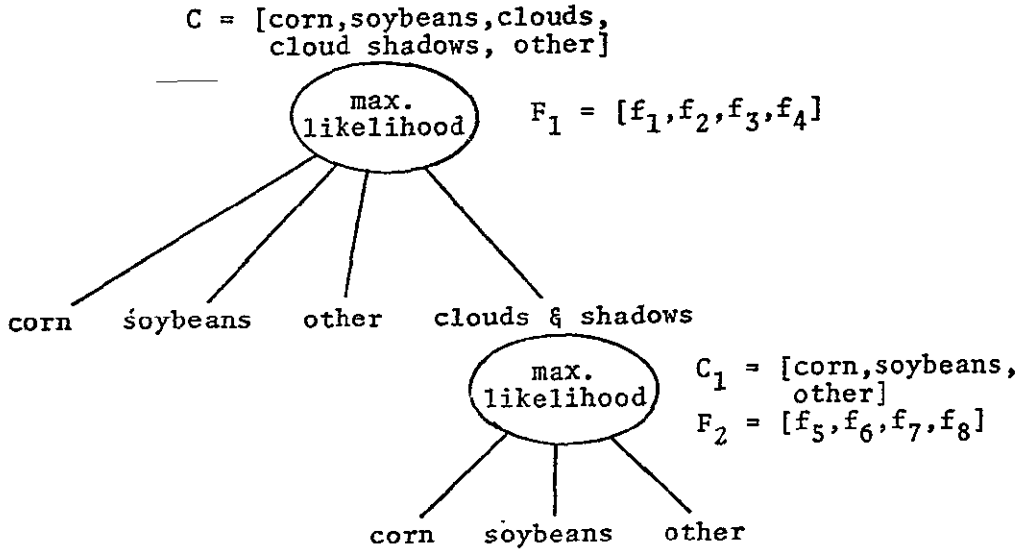


Figure 2.1-2 Agricultural classification in the presence of clouds.

-Spectral/geophysical data

*Detection of class specific properties

-Crop disease detection

-Forest type mapping

-Water temperature mapping

*Minimization of data dimensionality

-Efficiency (speed) improvement

-Minimization of the effects of severely limited training data

We have pursued the development of procedures for effective design and utilization of layered classifiers in order to exploit their potential for (a) dealing with classification problems not well suited to conventional classifiers, (b) improving the efficiency of the classification process, (c) improving the accuracy of the classification process.

Previous Work at LARS

The theoretical foundations for this classification approach and for the design of layered decision trees were developed at LARS in earlier contract years. A set of research software for both the design and implementation of layered decision logic was also developed. The final report for the immediately preceding contract year reported advances in the development of both manual and automated design procedures and the application of these procedures to a number of practical remote sensing problems. Solution of the multitemporal cloud problem described above was reported and the accuracy and efficiency of layered classifiers in remote sensing data analysis were demonstrated.

OBJECTIVES AND GOALS

The objectives of this research and development task are threefold:

1. To develop an effective method for the design and realization of layered classifier logic applicable to multispectral/multitemporal remote sensing problems;

2. To implement effectively the products of the development on a general purpose computer in order to make them available to applications-oriented research; and

3. To demonstrate the effectiveness of layered decision logic in the context of a number of current applications.

The goals of the research for this contract year focused on the development of layered classifier logic applicable to multitemporal data sets, particularly for the case in which the training data for all available channels cannot be drawn from a single set of points (a single training segment). Other goals included upgrading the research software for the layered classifier to make it LARSYS-compatible, and surveying possibilities for improving the procedures for designing optimal decision trees. The first of these goals was principally motivated by the potential applicability of the results to LACIE. The latter goals represent advancements in the ongoing development of data analysis techniques for remote sensing.

ACCOMPLISHMENTS

A candidate strategy was formulated for the multitemporal problem involving training data from multiple sites. Preliminary tests of this decision strategy were carried out using data from CITARS and LACIE, but tests adequate to provide a definitive evaluation are awaiting availability of suitable data sets.

The layered classifier approach was also applied to related problems, including change detection.

Software for implementing layered decision logic was upgraded considerably and made available to users at LARS. Compilation of user and programmer documentation for this software has yet to be completed but is well under way at this writing.

The process of designing optimal decision trees has been modeled as a four-stage process involving (1) feature selection, (2) cluster analysis, (3) specification of a classifier at each node of the tree, and (4) tree search. A research plan aimed at optimal choice of algorithms for each of these stages has been formulated and the work has been initiated.

DESCRIPTION OF RESEARCH

A. Applications of the Layered Classifier Approach

The Multitemporal/Multisegment Problem

One way to state the problem is as follows:

Given three segments (ground areas) A, B and C and two satellite passes over these segments, reference data is assumed available for segments A and B, but not for both segments at all times of interest. Segment C is to be classified using statistics from the available multipass data from segments A and B.

Assume that training data is available from segment A at time 1 (but not time 2) and from segment B at time 2 (but not time 1). The classification of segment C (data available at both times 1 and 2) can be carried out utilizing a sequential (layered) decision strategy.

We make three further assumptions:

1. Segments A, B and C belong to the same spectral stratum (i.e., similar ground covers in all three segments have similar spectral response characteristics);

2. The informational classes in segment C may be assumed to be a subset of the union of the classes in segments A and B;

3. No dramatic changes occur in climatic conditions during the acquisition of data which would cause the time 1 data for segment C to differ in spectral character from the time 1 data for segment A; and similarly for the time 2 data with regard to segments C and B.

The analysis strategy is as follows: Training is performed on the cloud-free passes of segment A and B utilizing the available ground truth. (Even if several cloud-free passes should be available, training is performed for each pass individually.) All spectral classes are investigated for separability and classes that are not separable are either deleted if determined to be unimportant or else combined, resulting in "mixture classes" in those cases in which the spectral classes were derived from different cover types. It is then decided which date (satellite pass) will serve as the base date to provide the initial stages of the classification logic. For classification of agricultural scenes, it is often most logical to use as the base date the data from the date at which the crops of interest are most discriminable. The stages of the decision tree for this data are then designed using the Optimal Decision Tree Design procedure. Terminal nodes of this tree which represent mixture classes are replaced by small subtrees which use data from a date other than the base date, if possible, to resolve the mixture into individual classes. The tree thus assembled is used in the classification of segment C.

There are a number of practical difficulties associated with this strategy, some of a minor technical nature, some more serious and fundamental. A minor problem is that the total number of spectral classes involved for a given case equals the sum of the number of classes obtained for each pass. This number can be very large and the software must be designed to handle both the volume and the considerable amount of "bookkeeping" involved with the classes and their associated statistics.

A more fundamental problem is that this procedure does not allow for explicit interaction among the data from different times. That is, the training statistics used at any decision node must be drawn from a single time/segment and any interactions between times which might improve the classification accuracy cannot be used to advantage. This restriction arises from the fact that a simple correspondence between classes in different segments (or even the same segment at different times) cannot be routinely established. We do not at this time have a practical solution to this problem implemented.

EXPERIMENTAL RESULTS

The data chosen for the analysis described below consisted of three multitemporal data sets assembled for LACIE. The data were collected during early 1974 over the Intensive Test Sites in Finney, Morton and Ellis Counties, Kansas. A map of the state of Kansas shows the position of the three counties relative to each other (Figure 2.1-3). The available dates and corresponding channel numbers in the overlaid data set are given in Table 2.1-1. Reference data for the three segments were available in the form of aircraft photography at a scale of 1:24000 taken in June 1974 and plastic overlays containing field boundaries and field numbers for each test site. A listing of the cover type for each field was available from field visitation reports. Unfortunately, the ground truth for Finney County was assembled prior to the time that the Intensive Test Site was moved southward because of an atypically large number of irrigated fields. Therefore, only ground observations about the southern part of the original test site could be used. It was decided that Finney and Ellis would be used as training segments "A" and "B," respectively, and Morton would serve as segment "C," the segment to be classified.

In order to ease the burden of manually picking training and test fields using the overlays, the three sites were geometrically corrected and scaled to 1:24000 for line printer output.

As information about spectral strata concerning these three sites was not available, grayscale prints and color composites were studied. The sites are approximately 130km (80mi) from each other and one might suspect that they would not belong to the same stratum. The color composites showed for all three available common dates that there were significant differences in spectral response. Channel 3 for each of the segments in each date is shown in Figure 2.1-4 (a-i). These also reveal large differences from segment to segment at each time. Channel means and variances are given in Table 2.1-2.

Based on these observations, it was clear that these three segments could not be considered to belong to a single spectral stratum, thereby violating one of the basic assumptions of our present approach. However, since these were the only data available to us with reference data and appropriately related with respect to timing of the respective satellite passes, we decided to press ahead with the analysis to see what could be determined.

Training and test fields were selected on a random basis for each segment, allocating roughly twice as many fields for testing as for training. To provide a basis for evaluating the layered classification results, each segment was classified by a single-stage classifier using training statistics from the same segment and pass (see Table 2.1-3a).

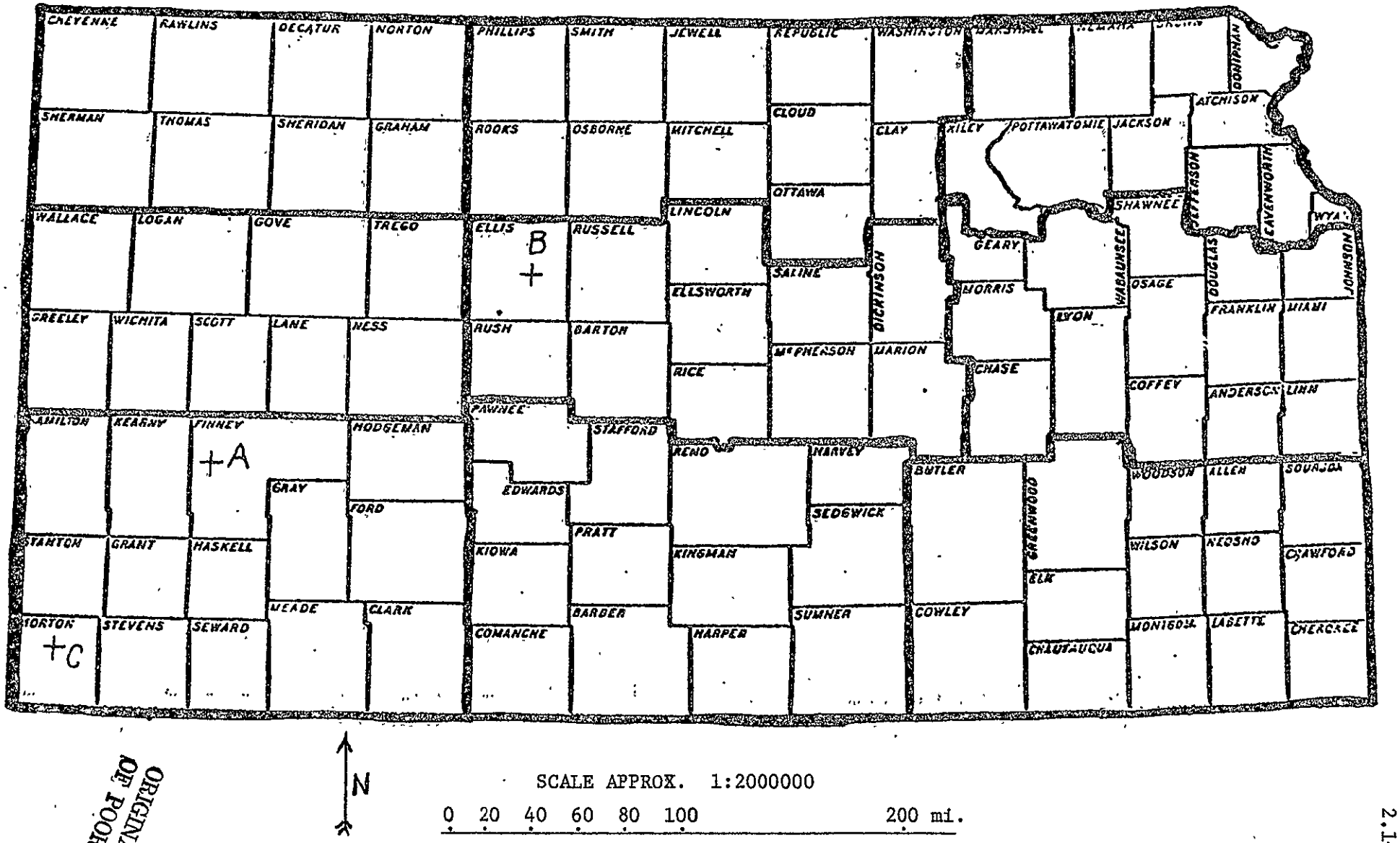


Figure 3 Test Site Locations

Table 2.7-1 Dates data taken for overlays used in multitemporal/multisegment problem

	FINNEY		MORTON		ELLIS	
Pass Designation	Tape Channel	date	Tape Channel	date	Tape Channel	date
			1 2 3 4	Oct. 23, 1973		
I	1 2 3 4	Feb. 25, 1974	5 6 7 8	Feb. 25, 1974	1 2 3 4	Feb. 24, 1974
II	5 6 7 8	May 26, 1974	9 10 11 12	May 27, 1974	5 6 7 8	May 26, 1974
					9 10 11 12	June 12, 1974
III	9 10 11 12	July 1, 1974	13 14 15 16	July 2, 1974	13 14 15 16	July 2, 1974

To estimate the impact of the three segments not belonging to the same spectral stratum, a single-stage classification was performed using the statistics from Finney and Ellis, separately, to classify the corresponding dates of Morton (Table 2.1-3b). The results showed marked deterioration of results as compared to Morton classified with Morton statistics.

In an attempt to ameliorate the multiple stratum effects, an additive shift was made in the training class means and the single-stage classifications of Morton were repeated. The results, shown in Table 2.1-3c, appeared sufficiently improved to warrant continuation with the experiment. However, we would not advocate this sort of treatment of the data on a routine basis. Nor does this provide a general justification for relaxation of the assumption that all segments be from the same spectral stratum.

The test of the multitemporal layered classifier logic consisted of using the mean-adjusted statistics from Finney II and Ellis III to classify Morton II + III. A decision tree was designed using Finney II as the base date and using Ellis III to resolve occurrences of the "mixture" classes. The results are shown in Table 2.1-3d.

To evaluate these results, one must first decide what to compare them against. There are a number of possibilities and we shall look at several.

1. We can compare the results with classifications of Morton using Morton statistics. Although in the practical problem, our hypothesis is that training data are not available from Morton, in the experiment this provides us a reasonable "upper bound" on what we might expect to achieve; i.e., this provides an indication of how separable the classes are in the classified segment.

2. We can compare the results with the classifications of Finney II using Finney II statistics and Ellis III using Ellis III statistics. The latter two classifications provide an indication of how separable the classes are in the training segments--another "upper bound."

3. We can compare the results with classifications of Morton II with Finney II statistics and Morton III with Ellis III statistics. This provides probably the best indication of the contribution of the layered classifier approach in the context of the actual application.

The reader who tries to make these comparisons based on the results in Table 2.1-3 will find it quite a task because of the erratic nature of some of those results (see especially Morton III classified with Morton III statistics). All of these analyses were performed using the same "standard" procedure, and we have not been able to attribute the peculiarities to anything specific except for some of the irregularities in the data set, especially the reference data. We look forward to conducting further tests on data sets of greater integrity.

Table 2.1-2 Channel Means and Variances for Finney, Morton and Ellis

	FINNEY			MORTON			ELLIS		
pass	ch	mean	var	ch	mean	var	ch	mean	var
				1	30.7	5.32			
				2	32.1	7.07			
				3	37.4	8.06			
				4	18.9	4.20			
	1	26.1	3.56	5	27.6	2.46	1	24.6	3.26
	2	27.0	4.62	6	29.4	3.71	2	24.7	3.56
I	3	28.8	5.02	7	33.0	4.82	3	26.9	4.81
	4	14.7	3.06	8	17.1	3.22	4	13.9	2.32
	5	35.9	7.11	9	46.8	6.89	5	30.8	7.58
	6	35.5	11.7	10	55.0	10.9	6	26.9	8.42
II	7	48.4	10.9	11	57.7	6.74	7	37.9	10.8
	8	24.5	7.74	12	27.2	3.56	8	19.8	7.37
							9	35.2	4.77
							10	33.5	7.91
							11	44.3	7.81
							12	21.8	5.04
	9	39.9	6.22	13	49.1	6.24	13	38.5	3.90
	10	40.3	12.7	14	56.5	10.1	14	40.8	7.02
III	11	57.7	8.34	15	59.7	10.5	15	48.0	7.65
	12	29.3	6.38	16	27.5	5.32	16	23.4	4.49

(a) Finney I

(b) Ellis I

(c) Morton I

(d) Finney II

(e) Ellis II

(f) Morton II

(g) Finney III

(h) Ellis III

(i) Morton III

Figure 2.1-4 Imagery of 3 sites, 3 passes (channel 3)

Table 2.1-3 Multitemporal/Multisegment Classifier Performance

Training Site	Classified Site	Wheat	Pasture	Other Grain	Fallow	
FI	FI	24.6	12.8	22.8	20.6	
FII	FII	58.2	58.6	25.8	23.3	
FIII	FIII	45.1	27.6	44.7	43.6	
EI	EI	42.3	49.3	-(note 1)	20.0	(a) Sites classified using "local" statistics
EII	EII	97.2	66.5	-	62.6	
EIII	EIII	89.7	10.5	-	68.2	
MI	MI	42.2	31.0	32.0	14.8	
MII	MII	79.3	52.3	16.9	-	
MIII	MIII	2.2	19.2	18.6	4.0	
FI	MI	14.5	5.9	16.5	23.1	(b) Morton classified using unadjusted statistics from Finney and Ellis (note 2)
FII	MII	9.7	4.9	89.8	18.8	
FIII	MIII	41.0	11.3	12.1	16.0	
EI	MI	38.7	25.6	-	11.4	
EII	MII	5.5	7.5	-	0.0	
EIII	MIII	97.6	0.0	-	5.1	
FI	MI	27.2	0.7	25.5	21.2	(c) Morton classified using adjusted statistics from Finney and Ellis
FII	MII	59.6	8.5	34.1	0.3	
FIII	MIII	23.1	15.5	38.5	0.8	
EI	MI	57.1	1.2	-	19.4	
EII	MII	82.0	0.0	-	41.8	
EIII	MIII	54.0	3.3	-	39.3	
FII+EIII	MII+MIII	62.9	4.9	33.2	10.3	(d) Multitemporal classification of Morton using Finney and Ellis statistics

note 1 - indicates no training data available.

note 2 - Ellis II cloud class had substantial impact.

The following observations from Table 2.1-3 are encouraging, however:

1. The multitemporal layered classification yielded slightly more accurate classification of wheat than either Morton II classified with Finney II statistics or Morton III classified with Ellis III statistics.

2. With the multitemporal data, it was possible to classify the "other grain" in Morton whereas this was not possible using the Ellis III statistics alone (Ellis III has no training data for "other grain").

Thus, although we cannot draw very strong positive conclusions from this single test, we feel that in the face of the data quality problem and the multiple stratum situation which had to be dealt with, the results are sufficiently encouraging to warrant further tests as appropriate LACIE data sets become available.

Change Detection

The layered decision strategy used for multitemporal classifications is easily adapted to change detection. Training is again performed separately for the respective passes. In the base date, all classes are considered, while in the second date the analyst need only account for the classes into which the change is assumed to happen. For the base date, a tree is designed using the Optimal Decision Tree Design procedure. A second tree is then designed using the second date for the possible change classes. The trees are joined such that the tree from the second date is attached to each terminal node of the base tree representing a class for which change is anticipated. Thus, data from the second date is used only for those points classified from the first date as candidates for change.

An experiment was performed using data over the Texas Coastal Zone consisting of a two date overlay, the first date November 27, 1972 and the second date December 15, 1973. Because of the relative dates of the two passes, seasonal differences were not expected to seriously influence the outcome of the change detection procedure. To keep the problem simple, only change from water to sandbank was investigated. Figure 2.1-5 shows the tree used in the procedure. Figure 2.1-6 shows part of the resulting classification. The shaded area shows points where water had changed to sandbank in the course of a calendar year. The result was compared to a "straightforward" change detection, in which the results of two classifications were compared on a point-by-point basis. The results of the two procedures were in good agreement. Based on the outcome of this experiment, further change detection work using the layered classifier was conducted under this contract as part of the Regional Applications Project.

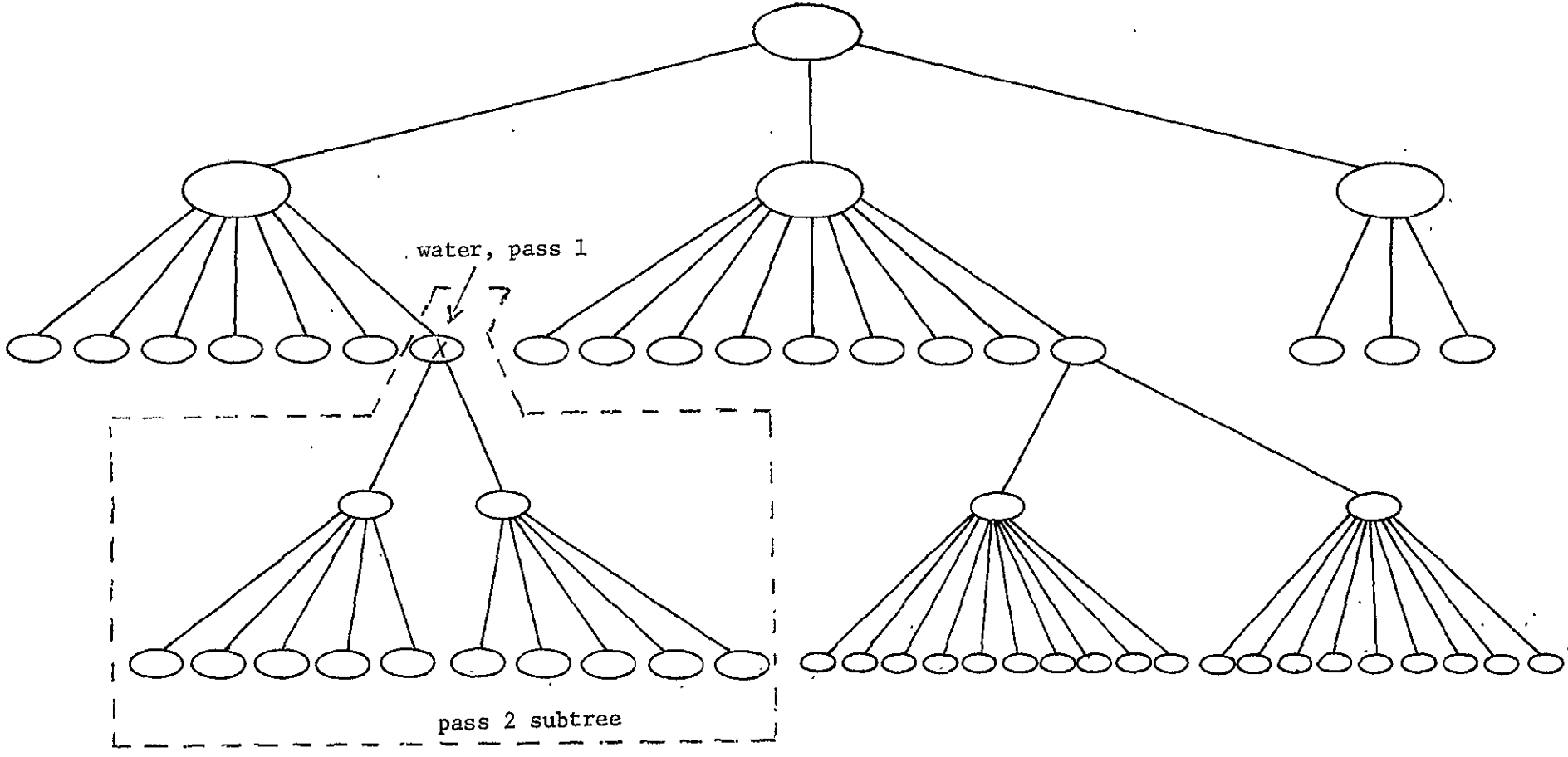
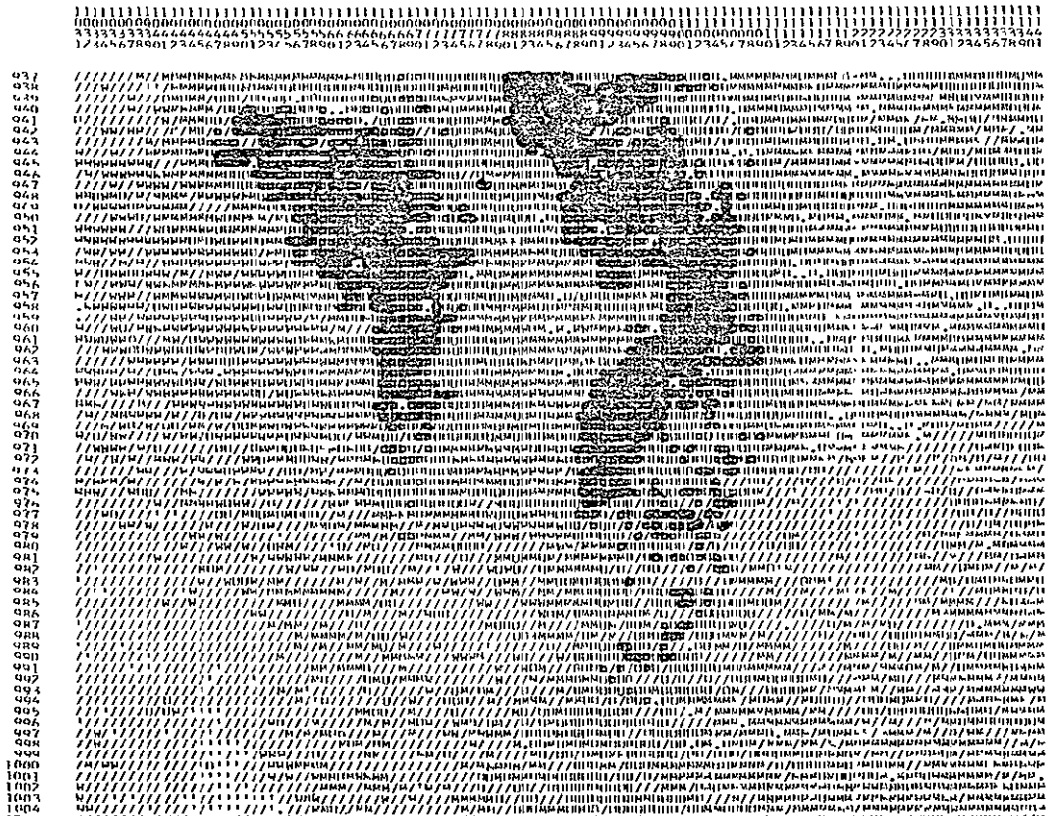


Figure 2.1-5 Decision Tree for Change Detection ("water" to "sandbank").



Darkened areas shown change from "water" to "sandbank".

Figure 2.1-6 Change Detection by Layered Classifier.

B. Layered Classifier Implementation

Approach

Results of previous research included computer programs which implement the layered classifier approach to data analysis in two phases. The first phase, called the Optimal Decision Tree Design procedure, is an algorithm for automatically determining the optimal layered decision logic for a layered classifier. The second phase accepts as input the description of layered decision logic, generated by the Optimal Decision Tree Design procedure or any other method, and performs classification of multispectral remote sensing data. We actually refer to this phase as the Layered Classifier.

The specific goal of the implementation effort is to upgrade the research software to the point of being LARSYS-compatible so that (1) this method of data analysis will be conveniently available to applications-oriented researchers who need it, (2) further upgrades to incorporate advances in the technology will be facilitated, and (3) the software will be sufficiently documented for transfer to NASA and others requesting it.

To accomplish this goal required detailed analysis of the existing software to determine its operation and the nature of the desired interfaces with both the user and other software needed in the analysis process; planning and execution of the software design; and generation of documentation appropriate both for users (data analysts) and programmers.

RESULTS

The first step taken was to go through the programs with an eye to how they should look when finished. A list of problems was laid out and ordered by priority. This list included the following:

1. What control cards are needed to input all necessary data?

2. In what format should the decision tree information be input?

3. What internal representation of the tree is best in terms of flexibility and storage efficiency?

4. How can all of this information be saved on a results tape for later display and further manipulation?

After examining the programs, it was decided that for expediency the LARSYS CLASSIFYPOINTS processor would be modified to perform layered classification. Figure 2.1-7 shows a diagram of the Layered Classifier which resulted.

The Layered Classifier research software used a fixed format data deck for the decision tree input. Because users could be expected to need to design and modify their tree decks manually, a free format tree representation scheme was necessary. This involved drastic changes to the format and necessitated the development of a new program utilizing original card reading and decoding techniques. Table 2.1-4 shows the control card description for the new layered classifier.

Internally, a new tree format had to be devised. The old one used fixed array sizes and greatly limited the number of nodes in the decision tree as well as the size of the data set to be classified. A new linked list system was developed to internally describe the tree. A dynamic memory allocation scheme was devised to increase the utilization of the available storage.

A method for graphically representing a tree on the computer output was developed. This should prove to be a great help to analysts as it provides a permanent, easily interpretable record of what their tree looks like. An example of this output is shown in Figure 2.1-8.

Some initial consideration has been given to the problem of putting the tree information on the results tape. Probably the easiest way is to copy the tree deck to the tape in card image form similar to the statistics deck. At present, however, this has not yet been implemented because of the impact upon other programs which read the results tape.

In summary, then, over the contract year a new Layered Classifier program was developed and implemented. This software was written according to LARSYS standards and is compatible with current LARSYS programs. The output tape is compatible with LARSYS PRINTRESULTS and

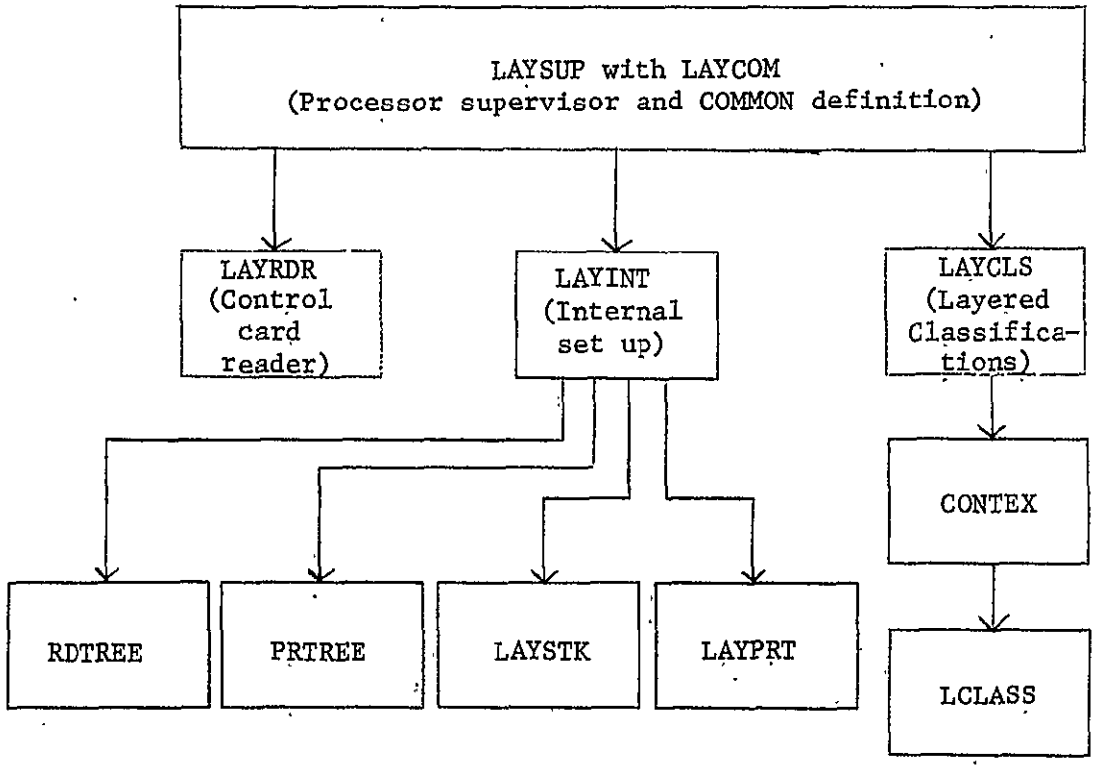


Figure 2.1-7 Layout of the Layered Classifier

Table 2.1-4 Control Card Summary for the Layered Classifier

REVISED 02/01/76

LARSYS CONTROL CARDS

LAYERED CLASSIFIER

REQ	KEY WORD(COL.1)	CONTROL PARAMETER	FUNCTION	DEFAULT
+	*LAYER	(NONE)	SELECT LAYER CLASSIFICATION FUNCTION.	(NONE)
+	RESULTS		DESTINATION OF RESULTS	(NONE)
		TAPE(XXX)	PUT ON TAPE XXX,	
		FILE(FF)	FILE FF.	
		INITIALIZE	INITIALIZE FILE ONE OF A NEW RESULTS TAPE (REQUIRED WHEN USING A NEW TAPE).	SEE CONTROL CARD DICTIONARY
		DISK	RESULTS WILL BE STORED ON LARSYS DISK.	
	PRINT	STATS	PRINT STATISTICS TO BE USED.	NO STATISTICS PRINTED
		MAP	PRINT RESULTS MAP.	NO MAP PRINTED
		NOFIELDS	NO TRAINING FIELDS PRINTED.	TRAINING FIELDS PRINTED
	CARDS	READSTATS	STATISTICS FILE WILL BE INPUT ON CARDS.	STATISTICS EXPECTED FROM DISK
	DATA	-----START OF DATA DECK----------		
		PUNCHED STATISTICS FILE FROM STATISTICS FUNCTION IF 'CARDS READSTATS' CONTROL CARD IS INCLUDED.		
+	DATA	-----START OF DATA DECK----------		
		DECISION DECK USED FOR THE LAYERED CLASSIFIER		
		CARDS ARE FREE FORMAT ON EACH CARD BUT CARD ORDER MUST BE PRESERVED. DECK IS ALWAYS REQUIRED.		
		DATA CARD FORMATS----		
		TREE TOP(NODEA,NODEB,NODEC),NODEA(NODE1,NODE3,NODE4) TREE NODEB(NODE2,NODE5), ETC		
		FEATURES TOP(X,Y),NODEA(Y,Z),NODEB(W,X,Y), ETC		
		REPRESENT NODE1(I,J),NODE2(K,L,M),NODE3(N),NODE4(O) REPRESENT NODE5(P), ETC		
		RENAME CLASS1(1/I,J/),CLASS2(2/K/),CLASS3(3/L,M/), ETC		
		THE TREE CARDS DESCRIBE THE NODE SEQUENCE OF THE DECISION. THE FEATURES CARDS DESCRIBE THE CHANNELS USED AT THE EACH DECISION NODE. THE REPRESENT CARDS ARE USED TO INDICATE THE REPRESENTATIVE CLASS USED. THE RENAME CARDS ARE USED THE ASSIGN A NEW NAME TO GROUPS OF INPUT CLASSES.		
		TOP IS A REQUIRED NAME ON THE TREE AND FEATURES CARDS. THE NODE AND CLASS NAMES MAY BE UP TO EIGHT CHARACTERS.		
+	DATA	-----START OF DATA DECK----------		
		FIELD DESCRIPTION CARDS DESCRIBING AREAS TO BE CLASSIFIED (ALWAYS REQUIRED). EITHER FORM OF THE FIELD DESCRIPTION CARD MAY BE USED.		
+	END	(NONE)	END OF FUNCTION.	(NONE)

LISTRESULTS. In addition, a set of software documentation and user documentation is being completed to describe the implementation and usage of the Layered Classifier. We did not reach the point of being able to expend effort on the Optimal Decision Tree Design software, however; it is desirable that this software be developed into a program similar to the Layered Classifier as it is an important component of the layered classifier system. Some additional effort should be directed toward appropriately saving the decision tree on the results tape together with the corresponding classification.

C. The Optimal Decision Tree Design Procedure

Approach

For a fixed set of features, classes, and candidate classification criteria (maximum likelihood is only one example of the latter), it is possible to enumerate all possible decision trees which could be constructed. In theory, one could evaluate each of these trees and select for implementation the tree which gives the best performance. In practice, however, the number of possible trees is generally so large as to preclude such an exhaustive search for the optimal decision tree, and we therefore seek to devise more clever means of determining the optimal decision tree.

In earlier research at LARS, Wu (Information Note 090174) developed a heuristic search procedure for decision tree design which we refer to as "guided search with forward pruning." Essentially, the strategy is to construct the decision tree a node at a time, estimating the suitability of all candidate substructures for the node under consideration, and discarding all but the most promising candidate subtree.

The strategy requires a means for evaluating each node in the tree. For each candidate structure following node d_i, the evaluation is computed as follows:

$$E(d_i) = -T(d_i) - K \cdot \varepsilon(d_i) + \sum_{j=1}^{n_i} E(d_{i+j})$$

where $T(d_i)$ and $\varepsilon(d_i)$ express the efficiency (computation time) and accuracy associated with node d_i and the summation is an estimate of the evaluation functions of the descendent nodes of d_i (assumed to be n_i in number). A lower bound is used to form this estimate, which is the evaluation of a conventional single-stage classifier applied at that point.

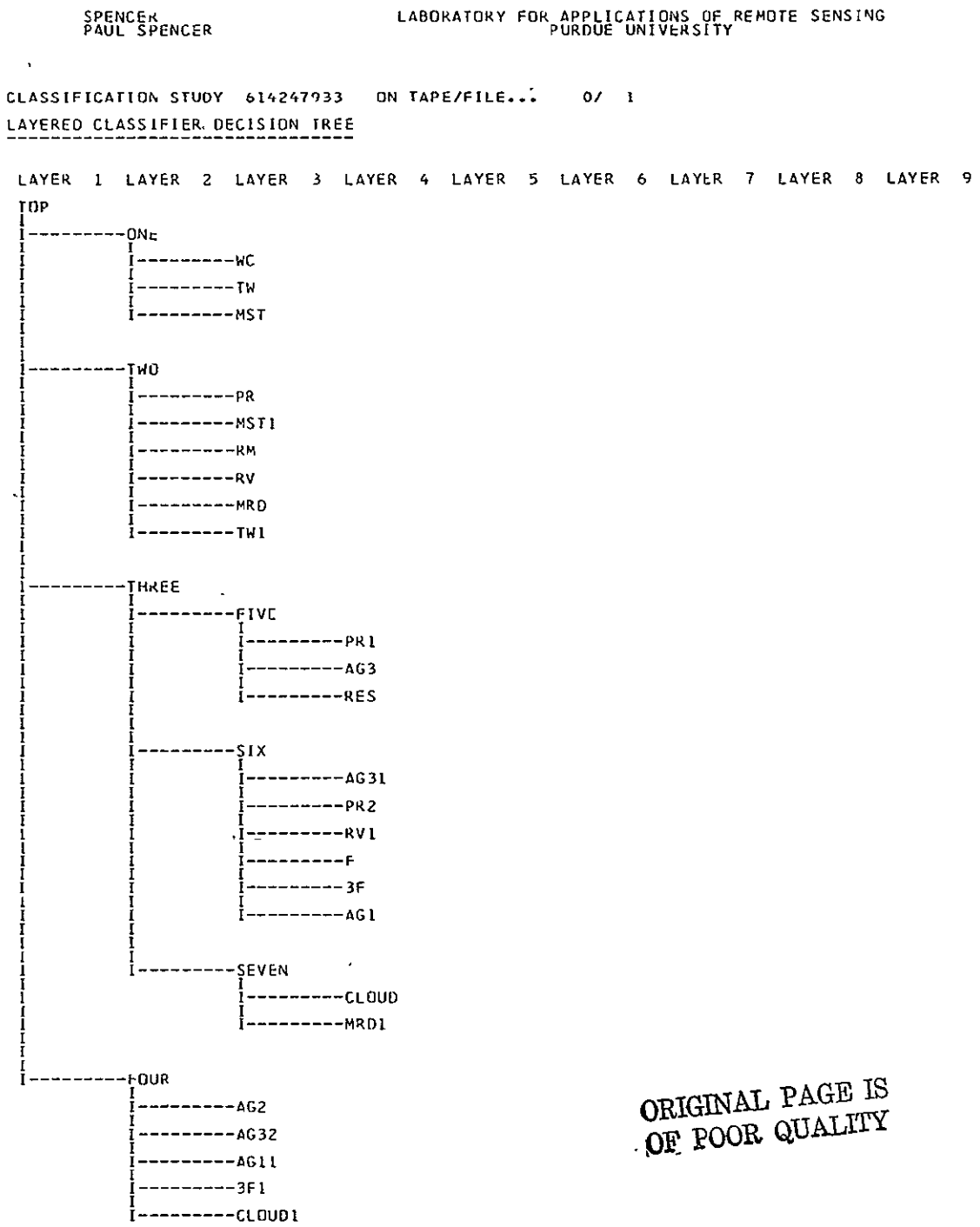


Figure 2.1-8 Example of Graphical Display of the Decision Tree.

Constructing the tree in this sequential fashion cannot guarantee that the optimal tree will be obtained, because, unfortunately, the optimal choice at any level in the tree is not necessarily independent of choices at later stages. Still, the suboptimal results have been demonstrated in many applications to be an improvement over the alternative single-stage classifiers. Research is proceeding in the direction of improved heuristic search and mathematical programming techniques to further improve the decision tree design.

RESULTS

Four key steps in the design process have been identified: (1) class clustering at each node, (2) feature selection at each node, (3) specification of the classifier at each node, and (4) tree search. The first of these steps partitions the set of classes into the class subsets to be discriminated at the node; the second assigns the appropriate subset of the available features to be used to discriminate the class subsets; the third step determines the degree of complexity to be implemented in the decision process at the node; and the fourth step implements the search through all candidate trees for the desired (and, ideally, optimal) decision tree.

A research plan has been formulated and is being pursued which involves the systematic consideration of several alternative methods which could be applied to each of these key steps in the decision tree design procedure. However, further results are not yet available.

RECOMMENDATIONS

Research in the design and application of layered classifiers has demonstrated the great potential of this generalized form of data analysis by pattern recognition. Continuation of this research can be expected to be very productive. In particular, we recommend that the following be pursued vigorously.

*Applications in the temporal domain. Continue to develop and improve layered decision logic for a range of multitemporal/multisite applications. Demonstrate the performance achievable by acquiring and analyzing appropriate data sets. Determine the impact of the spectral stratification research on the layered classification approach.

*Software upgrade and documentation. Complete documentation of the *LAYER (Layered Classifier) processor. Perform similar upgrade of the software for Optimal Decision Tree Design in order to facilitate incorporation of new research developments, make the capability available to applications research, and facilitate transfer of the software to other potential users, especially NASA.

*Optimal Decision Tree Design research. Pursue techniques for determining the optimal layered classifier logic. Much remains to be done in this area, and the potential exists for improving both the accuracy and the efficiency of the classifiers produced for any given application.

ACKNOWLEDGEMENTS

The project scientist for this task was Dr. Philip H. Swain, School of Electrical Engineering, Purdue University. Major contributions to the research and development were made by Dr. Hans Hauska, ESRO Research Fellow, at Purdue University on leave from Lulea University of Technology, Lulea, Sweden; and by Paul Spencer, Applications Programmer, and Sidney McAhren, Senior Student Programmer, both in the Computer Facility at LARS.

2.2 Development of Signature Extension Strata from Clustering Techniques

INTRODUCTION

Full realization of the potential advantages of the synoptic coverage provided by LANDSAT will require the development and use of data analysis techniques which take into account the large variation and diversity of patterns found over many LANDSAT scenes. Analysis techniques which are satisfactory for data acquired by airborne sensor systems or for limited areas of LANDSAT data cannot be effectively used to classify an entire LANDSAT frame of data. Fortunately, however, the variation found in LANDSAT scenes is not random, but occurs in very definite patterns. These landscape patterns are associated with the different topographic features, soils, crops, farming practices, and climatic zones found in a 35,000 square Km area.

This suggests that one of the first steps in the analysis and classification of LANDSAT data covering one or more LANDSAT scenes is to divide the scene into areas that have similar characteristics. Division of a heterogeneous population (or area) into subpopulations (or subareas), each of which is internally homogeneous is known as *stratification* and is frequently used by statisticians performing surveys to increase the precision of estimates.[1] If each stratum is homogeneous in that the measurements vary little from one unit to another, a precise estimate of any stratum mean can be obtained from a small sample in that stratum. Estimates from several strata can then be combined into a precise estimate for the whole population. Use of stratification in the sampling designs used for remote sensing applications is therefore advantageous.

A second use of stratification directly related to remote sensing applications is to potentially permit training statistics developed for one segment or portion of the scene to be successfully used to classify other segments which are spatially and/or temporally removed from the training segment. In this context the term *spectral stratification* is useful in that it connotes the division of the scene into areas which are internally spectrally similar. A spectral stratum may be defined as an area within which the scene (including atmospheric effects) is sufficiently similar that training statistics from one portion of the stratum can be used to classify other portions of the stratum without significant change in classification performance. Conversely, if the same training statistics are applied to areas outside the stratum in which they were developed, classification performance will decrease.

Computer-implemented clustering techniques provide an objective and efficient method for determining the similarity of units within LANDSAT scenes. The objectives of this project are: (1) develop multivariate pattern recognition procedures for determining and delineating spectral strata in LANDSAT data and (2) determine quantitatively the physical factors which account for the spectral strata.

Clustering Procedures for Spectral Stratification

The technique of clustering has been adopted to define the spectral strata present in LANDSAT scenes. Clustering has been used extensively in remote sensing to group together units which are similar, based on observation vectors and a measure of similarity. Most remote sensing data analysts are familiar with the process of clustering pixels into spectral classes to be used later in classification. The observation vector in that type of clustering is the spectral response of the pixel in each waveband, and a commonly used measure of similarity is the Euclidean distance in the observation space.

In spectral stratification, the sample unit is much larger than a single pixel and the objectives of the clustering technique are slightly different from the familiar process mentioned above. Instead of grouping together vectors of spectral responses for single pixels, we wish to group distributions of the spectral responses of sample units. Two units are spectrally similar if the distribution of spectral response in one unit is close to the distribution of the spectral response in the second area.

We can state the generalized procedure for clustering to define spectral strata in five steps.

1. Select sample units in the scene.
2. Characterize the distribution of the spectral response of each unit.
3. Choose a measure of similarity.
4. Apply a clustering algorithm to the units to determine groups of spectrally similar units.
5. Delineate the strata boundaries.

Each step and its application to stratification will be explained further.

Selection of Sample Units

The sample units to be used in this procedure may either be segments whose geographic position has been fixed by a sampling scheme before the LANDSAT data is acquired or rectangular areas chosen from the LANDSAT data itself without regard for their geographic position. The size of the sampling unit affects the kind of strata that can be found as it is the effective lower limit on the size of strata that can be observed. For example, if the sampling unit is larger than the largest city in the scene, then urban areas cannot be separated as distinct strata. The smaller the sampling unit chosen, the smaller the geographic extent of the strata and the finer the division, or levels, that can be observed. For example,

if a pixel is chosen as the sampling unit, the strata essentially are the spectral subclasses of cover types present in the scene.

Characterization of Spectral Response

The distribution of the spectral response within a sampling unit may be characterized in several ways. Two methods have been pursued in this work. In the first method the distribution of the spectral response in an area is represented by its first and second moments, that is, by its mean vector and covariance matrix. These parameters are easy to calculate and to use with similarity measures. However, they do not contain complete information on the skewness, multimodality, and non-normality of the distribution, all of which may be important in applying a statistical measure of distance between distributions.

A second method is essentially non-parametric. The distribution of the spectral response is characterized by the marginal density functions of the distribution. The marginal density functions rather than the joint density function are used to meet computer space limitations when dealing with large numbers of sample units. The characterization of distributions of the sample units is accomplished by first tabulating a base histogram for each feature (wavelength band) for the entire scene which is to be stratified. Equally probable bins are established from these histograms. Then a vector is constructed for each sampling unit in which each entry in the vector is the number of pixels in the sampling unit which fall in the corresponding bin in the base histograms. Thus the histograms or marginal densities of each sampling unit are characterized relative to the base histograms. The "histogram vectors" formed in this manner can then be used as data by a clustering routine.

Similarity Measures

In addition to the choice of characterization of the distribution of each unit's spectral response, a choice must be made of how to measure the similarity of two or more sample units. Sample units will be spectrally similar if the distance between their distributions or density functions is small. For the first method, that of representation of a distribution by its mean vector and covariance matrix, several statistical measures are possible.[2]

The transformed divergence has been the primary similarity measure used in this research as its properties are closer to the Jeffreys-Matusita distance[2] than are the properties of divergence, yet it is computationally less complex that the Jeffreys-Matusita distance. The desirable properties of the Jeffreys-Matusita distance are that it is a metric among multivariate normal densities and it is related to the probability of error (amount of overlap) between two densities.

The implementation of these distance measures assumes that the distributions involved are multivariate normal. The assumption of

normality may be violated when the sampling unit contains bad data or clouds which saturate the dynamic range of the data or when the sampling unit is divided into two distinct spectral classes, leading to bimodality. Use of large sample units has tended to alleviate the second problem, and we have tried to avoid bad data lines. Examinations of histograms have indicated that the normality assumption is not unreasonable for the data we have been using.

For the second method, that of "histogram vectors", the Euclidean distance between the vectors was chosen as a similarity measure for two reasons. First, it is a familiar measure whose properties are well known, and secondly, it has been previously implemented and extensively used in clustering analysis.

Clustering of Sample Units

Once a characterization of the spectral response and a distance or similarity measure have been selected, groups of spectrally similar units must be determined. If the analyst were to manually examine all possible pairs of units, the process would quickly become unwieldly and the results difficult to interpret for a large number of units. For example, if 150 units are to be stratified, over 10,000 pairwise comparisons are necessary. A machine-implemented clustering algorithm calculates the many pairwise distances and combines the information before presenting the analysts with the natural groups of the sample units.

Two clustering algorithms have been applied in this project. The first is an iterative algorithm which has been available for both observation space and parameter space clustering.[2] The algorithm can be simple stated in its general form.

1. Determine initial group centers.
2. Assign each unit to the nearest group center.
3. If no unit has changed allegiance, go to step 4. Otherwise, determine new group centers and return to step 2.
4. If groups are distinct, stop. Otherwise modify the number of groups, determine new group centers, and return to step 2.

In our research this algorithm has been applied to cluster units characterized by their mean vectors and covariance matrices in the parameter space, and to cluster the histogram vectors in the observation space manner.

The second clustering algorithm is a systematic procedure for grouping spectrally similar units in such a way as to minimize the total

number of groups while avoiding the grouping of non-similar units.[3] This procedure is slightly more complex than the first, as is seen in the following description.

1. Assign each unit to its own group, G_1, G_2,..., G_n.

2. Order the pairwise distances (d_{ij}), by magnitude. The algorithm considers (d_{ij}) in ascending order. Let d_{xy} equal the smallest d_{ij}.

3. If $d_{xy} > T$, a threshold of non-similarity, grouping is completed. Otherwise, proceed to step 4.

4. If the units x and y belong to the same group, go to step 7. Otherwise proceed to step 5.

5. Construct the average distance $\bar{d}_{xu}$ between G_x and each other group $G_x \neq G_u$ for which $d_{ab} < T$ for all a in G_x and b in G_u. The average distance between groups is defined as the average of all pairwise distances between units in the different groups.

6. If $\bar{d}_{xy}$ is the minimum of the set of inter-group distances constructed step 5, then combine G_x and G_y into one group.

7. Set d_{xy} to the next d_{ij} and return to step 3.

We have used this algorithm to group the spectrally similar units characterized by their mean vectors and covariance matrices.

Delineation of Strata Boundaries

After clustering is completed, the strata boundaries are delineated. Presently, this process is done manually when full LANDSAT frames or portions of frames have been stratified, although in the future we intend to adapt the "Extraction and Classification of Homogeneous Objects" (ECHO) approach to establish the boundaries of strata determined on the basis of fixed segments or a small sample of a LANDSAT frame.[4] When fixed segments based on a sampling scheme are stratified, a list of the segments in each stratum is produced rather than a map since this is the knowledge desired in this case and since the geographic location of strata boundaries between the segments is uncertain. That is, even though it is known that the boundary is between certain segments, the exact location is unknown.

Data Set Available for Stratification

In the course of this contract, two data sets have been supplied to LARS by NASA/JSC. The first data set (LACIE) was received in July, 1975, and contained LANDSAT-1 data covering twenty-five 1973-74 LACIE segments,

including five Intensive Test Sites. All of these images were 117 lines by 196 pixels. Only the Intensive Test Sites were multitemporal. Training field coordinates and identifications made by photointerpretation were available for all segments and test field coordinates and ground-level identifications were also available for the Intensive Test Sites.

The second data set (SRS) was received at LARS in August, 1975 and contained multitemporal images of LANDSAT-1 data covering eighteen SRS sites and five Intensive Test Sites. The SRS segments were all 200 lines by 200 pixels in size while the Intensive Test Sites were larger and varied in size.

In addition, several full-frames of LANDSAT-1 data for Kansas acquired during the 1973-74 winter wheat growing season were available in-house at LARS.

Stratification of Full-Frame Images

Several full-frames of LANDSAT-1 data were stratified under this effort. During preliminary work, scenes 1583-16525 (Feb. 26, 1974) and 1689-16382 (June 12, 1974) were stratified using a sample unit of 200 lines x 200 pixels or 13 x 16 Km. The distributions of the spectral response of the sample units were characterized by both methods described in the section Characterization of Spectral Response, and both algorithms described in the section Clustering of Sample Units, were used to define spectral strata within the above scenes.

These preliminary stratifications were quite "blocky" and seemed to correlate only with major soil or land use divisions. This suggested that the size of the sample unit should be decreased, and so sample units of 100 lines x 100 pixels (or 6.5 x 8 Km) and 50 lines x 50 pixels (3.2 x 4 Km) were used in further investigations of full-frame stratification.

Sixteen LANDSAT-1 scenes over Kansas acquired during the 1973-74 winter wheat growing season and one LANDSAT-2 scene acquired in 1975 were stratified using a smaller sample unit size. Generally, strata could be found which correlated with soil and land use patterns. However, in these stratifications, there were sample units which were not geographically linked to other units in the same stratum. That is, sample units from one stratum were surrounded by units from another stratum. When these sample units were examined closely, it was discovered that they did have spectral characteristics which distinguished them from the surrounding units and allied them to units which were spatially removed. This result, though perhaps counter to an intuitive conception of strata as spatially contiguous, is consistent with spectral strata as regions which are internally spectrally similar.

The accuracy of the stratification can be assessed indirectly by comparing the strata found by clustering with maps of physical factors which are known to influence spectral response. Presently the LANDSAT

imagery, strata maps, and physical factor maps are being compared manually. The illustrations in Figures 2.2-3 permit a qualitative comparison of a spectral stratification and soil and land use maps for the same area of southwestern Kansas. The spectral stratification shown in Figure 2.2-1 was produced by the "histogram vectors" method described in the section on Clustering Procedures for Spectral Stratification. The sample units in this example are 50 pixels x 50 pixels or roughly 4 Km x 3.2 Km.

The soil association map shown in Figure 2.2-2 exhibits several features easily seen in both LANDSAT imagery of the area and the spectral stratification. The areas of the Udic Ustolls (12) are easily visible, as are the patterns of the Typic Ustolls (9, 10, and 11).[5] The land use map, Figure 2.2-3, was developed from LANDSAT imagery acquired during June and July 1973.[6] Almost two years later, the same land use patterns can be found in the stratification of the May 21, 1975 image.

These evaluations of full-frame stratification are subjective. A complete, more objective comparison of the stratifications with physical factor ancillary data was planned. But due to problems in the registration of ancillary data (see section on Ancillary Data Registration) and redirection of stratification efforts to the partitioning of LACIE segments (see section on Partitioning of LACIE Segments), this evaluation has not been carried out.

Later, when the digitization of the ancillary data is completed, we plan to conduct a regression analysis which will quantify the degree of correlation between the strata and various physical factors. Such an analysis will not only provide a measure of the accuracy of stratification, but also provide quantitative information on the influence of major agronomic and meteorological factors on spectral reflectance. The physical factors being investigated include crop maturity stage, soil association, land use, precipitation, temperature, and grain yield.

Partitioning of LACIE Segments

At the end of November, 1975, NASA/JSC requested that all segments in the two data sets which had been supplied by JSC be stratified, or partitioned into strata.

To accomplish this, the segments in the above data sets were separated into groups according to crop reporting district and wheat biophase. The majority of the segments in the two data sets lie in the Central and the Southwest Crop Reporting Districts. Of the remaining crop reporting districts in the western two-thirds of Kansas, only the South Central Crop Reporting District had sufficient segments in a biophase to partition within itself. In order to stratify all the segments available to us, an additional partitioning effort was made in which the segments outside the Central and the Southwest Crop Reporting Districts were assigned to the closest of those two major crop reporting districts. These extended groups were then partitioned.

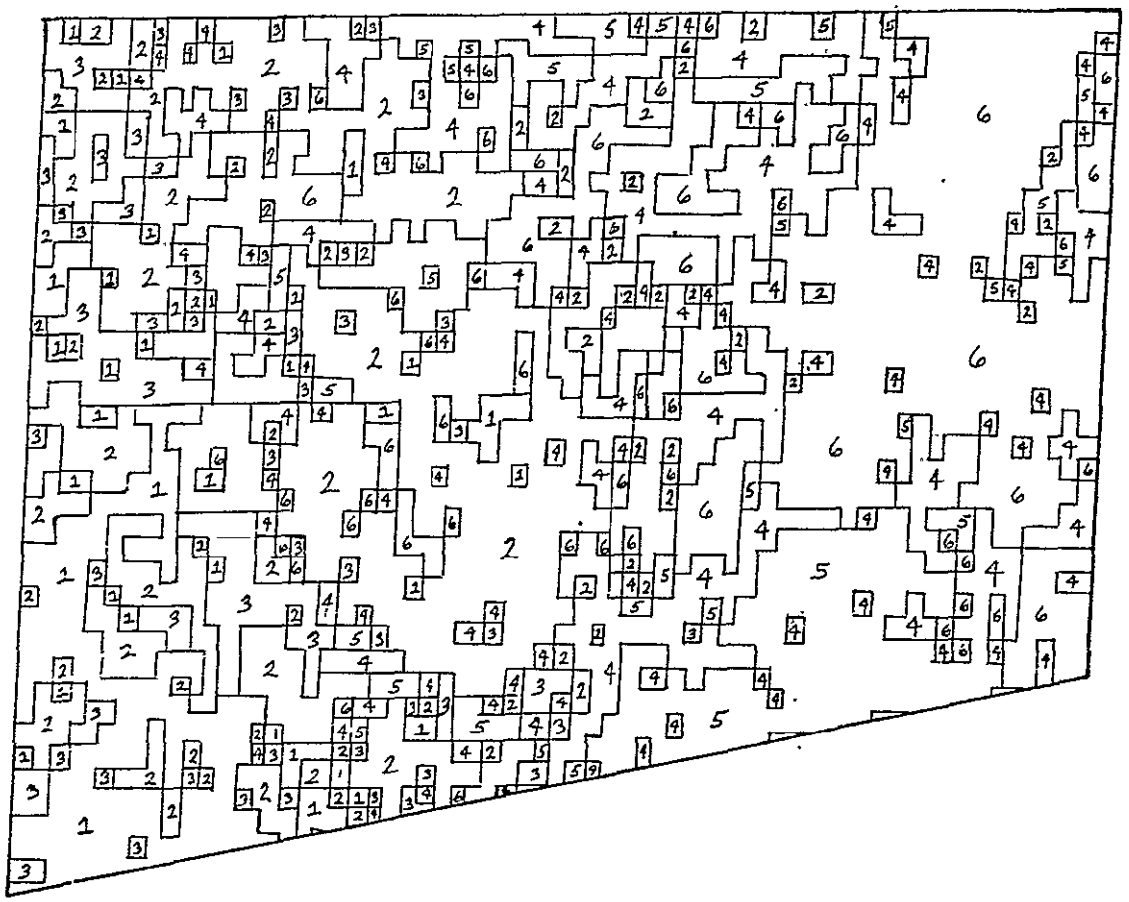


Figure 2.2-1

Machine-Implemented Stratification of the Kansas Portion of LANDSAT Scene 5032-16310. Each number represents a different stratum.

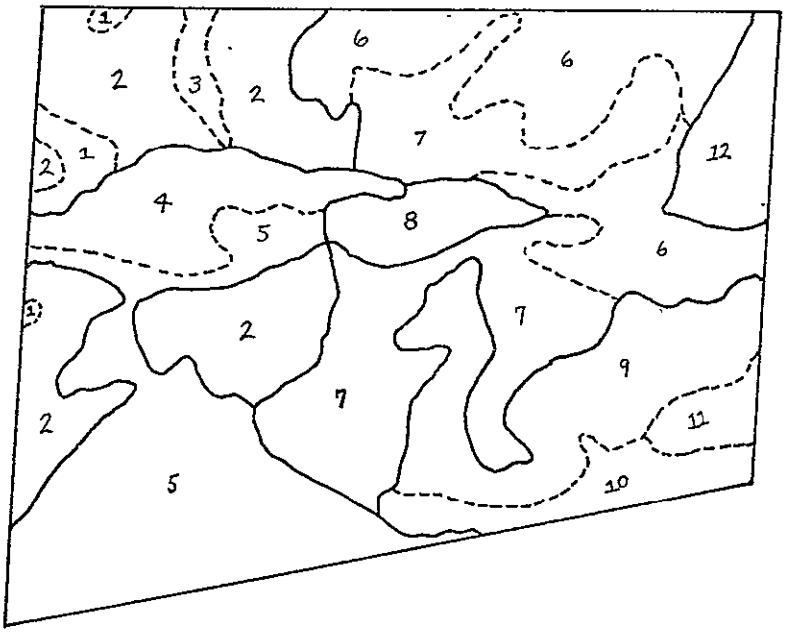


Figure 2.2-2. Soil Association Map of the Kansas Portion of LANDSAT Scene 5032-16310.

SOILS ASSOCIATIONS

ARIDIC USTOLLS
Ustolls, Orthents, and Ustalfs
Deep, grayish-brown and dark grayish-brown silt loams
1. Ulysses, Colby
2. Richfield, Ulysses
3. Ulysses, Drummond
Ustalfs, Psamments, Ustolls, and Argids
Deep, grayish-brown silt loams and sandy loams, and pale-brown loamy fine sands and fine sands
4. Tivoli, Vona
5. Dalhart, Richfield, Vona
TYPIC USTOLLS
Ustolls and Usterts
Deep and moderately deep, dark grayish-brown silt loams and moderately deep, gray clays
6. Harney, Uly, Wakeen
7. Harney, Spearville
Ochrepts, Ustolls, Ustalfs, and Psamment
Moderately deep and shallow, reddish-brown loams and clays, and deep, grayish-brown silt loams and clay loams and pale-brown loamy fine sands and fine sands
8. Manter, Pratt
9. Mansic, Mansker
10. Tivoli, Pratt
11. Woodward, Carey
UDIC USTOLLS
Ustalfs, Ustolls, and Aquolls
Deep, dark grayish-brown loams and fine sandy loams and pale-brown loamy fine sands
12. Pratt, Carwile

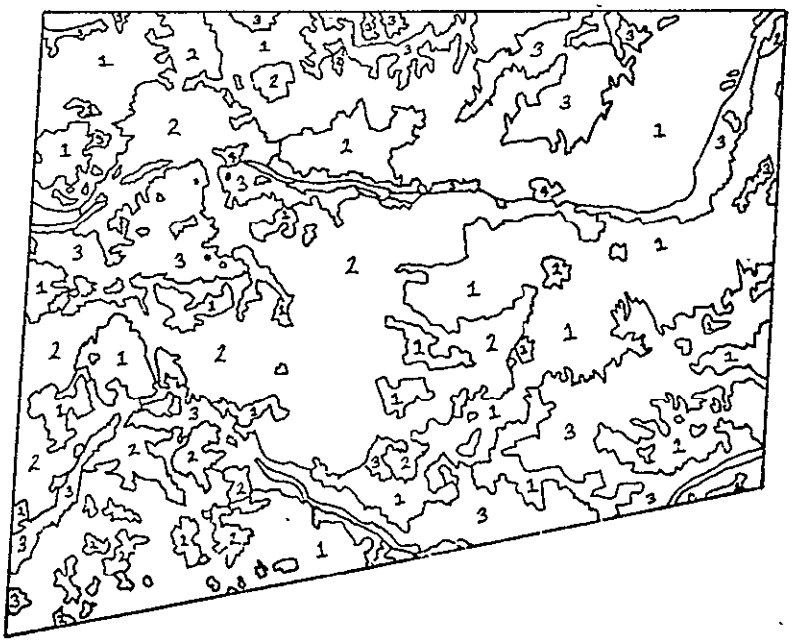


Figure 2.2-3. Map showing Major Land Use Categories for the Kansas Portion of LANDSAT Scene 5032-16310.

LAND USE CATEGORIES

1. Unirrigated - areas with greater than 50% unirrigated cropland
2. Irrigated - areas with greater than 50% irrigated cropland
3. Rangeland - areas with greater than 50% rangeland
4. Urban and built-up land
5. Water and wetlands

The four procedures used for this stratification effort can be easily specified in terms of the general procedure given in the section on Clustering Procedures for Spectral Stratification. In Step 1 of the generalized procedure, the sample units were chosen in two ways. First, each segment was considered one unit, except for the larger intensive test sites which were divided into units approximately the size of the other segments. Thus, the sample units were between 117 lines x 196 pixels to 200 lines x 200 pixels in size.

The second sample unit size chosen was approximately 100 lines x 100 pixels. In this selection of sample unit, a sample unit was a quarter-segment, and sample unit size ranged from 57 lines x 98 pixels to 100 lines x 100 pixels.

In either set of sample units, the spectral response of the unit was characterized by the mean vector and covariance matrix of the pixels in the sample unit. The similarity of two sample units was measured by transformed divergence. Both clustering procedures, given in the section on Clustering of Sample Units, were applied to the groups of sample units of each size, and lists of segments in each stratum produced. Thus, four partitions were generated for each group of segments within a crop reporting district and wheat biophase.

Generally, the stratifications produced by the two clustering procedures were consistent for each sample unit size. The first algorithm usually generated fewer strata than the second clustering algorithm, and these strata were combinations of the strata produced by the second algorithm. When the smaller sample unit was used, parts of the LACIE segments were often assigned to different strata. The distributions of the smaller sample units are more prone to problems of nonnormality, skewness, and multimodality which may contribute to this effect.

We have attempted to measure the success of stratification of LANDSAT data by clustering through classification performance in local and non-local recognition of LACIE segments. The criterion for success is that classification accuracies for all segments within a stratum classified using training statistics developed within a given stratum should be similar.

To statistically evaluate a stratification, two or more areas with known crop identification data must be available within each stratum. These test areas should fall entirely within the stratum, and should be large enough to conduct a reasonable classification analysis. Such a data set will give an adequate test of the stratification of the test areas, but cannot be used to determine the accuracy of the strata boundaries. The data sets presently available to us do not meet these requirements and a statistical evaluation of the classification evaluation is not possible. A portion of the classification results are presented in order to draw some limited conclusions.

Classification results from one stratification of segments in LANDSAT scenes acquired June 12, 1974 for central Kansas are presented in Table 2.2-1. Each segment is either a 8x9.7 or 4.8x4.8 Km area for which the crop types and LANDSAT data coordinates of the agricultural fields are known. The stratification procedure treated the segments as the sample units and characterized each segment by its first and second moments. The procedure placed the segments from Stafford, Ellis, Ellsworth, and Rice Counties in one stratum along with one of the segments from Barton County. The other segment from Barton County was placed in a different stratum. Both of the procedures, described in the section on Clustering of Sample Units, gave the same result when transformed divergence was used as the similarity measure.

The classification results show that the stratification technique was successful in identifying segments which are indeed different. In no case was a high classification performance achieved when using training statistics from segments outside the stratum. For segments identified as members of the same stratum, similar high (approximately 90 percent correct) classification performances were obtained for both local and non-local classifications of several combinations of segments. This indicates that these segments are from the same stratum. But, in several other instances the non-local classification result was lower than the local classification performance, indicating these segments may be from different strata. This would mean that the clustering procedure is grouping the segments into groups or strata which are too broad.

Similar results have been obtained with two other crop reporting district-wheat biophase combinations. With the available data, however, we cannot state with certainty whether the stratification procedure should be modified or whether the inconsistencies in results are due to limitations of the available data sets. The evaluation of stratification procedures by classification performance has been limited by the resources available and plagued by problems with the data sets such as misregistration of multitemporal data, mislabelled fields, and inadequate amounts of training data. Lack of a more adequate test data set is a major problem at this time; greater emphasis will need to be placed on this requirement of stratification evaluation before additional progress can be made.

Ancillary Data Registration

The requirement for registration of ancillary data (soils, meteorology, etc.) is based on the requirement to relate spectral strata to physical and environmental factors in a quantitative manner. The assumption is that the variations in spectral response for a particular crop is related to the crop characteristics and the physical environment and knowing this relationship will enable extension of crop recognition algorithm parameters from one location to a distant location.

Table 2.2-1. Classification Performances (Wheat vs. Other) for Segments Within and Outside of Strata Determined by Clustering

Strata No.	Source of Training Statistics	Areas Classified*					
		Barton-1	Barton-2	Rice	Ellsworth[1]	Ellis	Stafford
		Overall Percent Correct					
1	Barton-1	83.7	42.9	15.1	69.4	54.1	61.5
2	Barton-2	27.1	96.0	93.8	90.0	56.2	52.5
2	Rice	34.1	92.0	93.4	85.7	47.4	69.1
2	Ellis	63.4	43.4	26.4	60.4	64.8	51.4
2	Stafford	58.2	55.4	42.0	59.9	61.7	89.9

* LANDSAT scenes 1689-16392 and 1689-16385 acquired June 12, 1974 over Central Kansas.

[1] Ellsworth was not used as a source of training statistics because only wheat field coordinates were available.

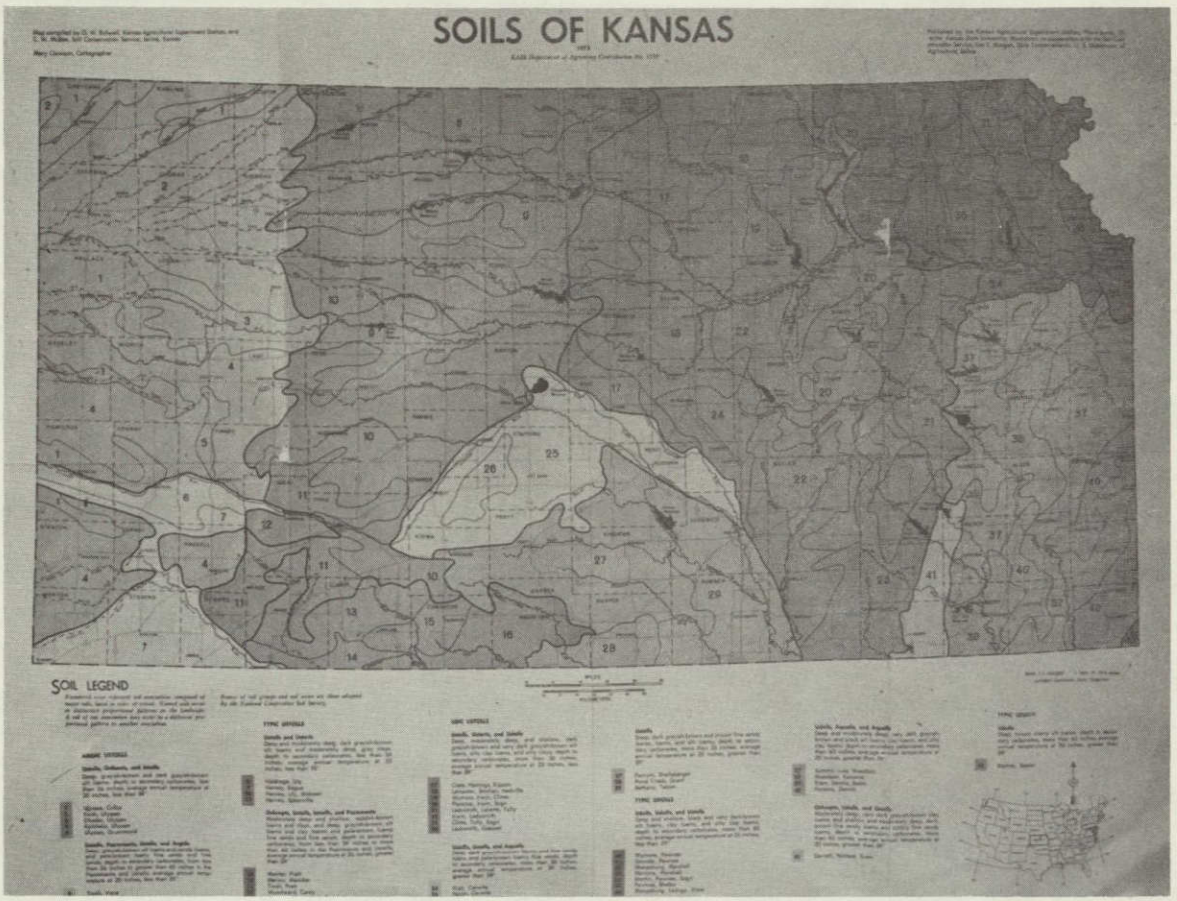


Figure 2.2-4. Source Map for Soil Association Data.

The ancillary data registration activity during CY76 was successful in achieving digital registration of soil, land use, and temperature data for one frame of LANDSAT data. Although the goal was to register several frames in this period, data processing problems were numerous and difficult, and achievement of the single frame registration is considered to be a significant milestone. The digital channels of physical factor data will enable correlation and other forms of numerical analysis to be performed on the physical factors rather than using visual interpretative means to related the factors to spectral groupings or strata.

The four variables registered for frame 1689-16382 acquired June 12, 1974 are soil association, land use, and precipitation for two time periods. The details of the registration processing are described in the section Research in Remote Sensing Technology. The source data and registration results are discussed here.

Soil Map Registration

The source soil map contains 42 soil units and was produced at a scale of 1:1,125,000.[5] A Lambert conformal conic projection was used for the map which is reproduced in Figure 2.2-4. The map was mounted on a coordinate digitizing table and each soil boundary was manually traced and digitized. Coordinates punched on cards were processed by a boundary data registration system to produce an additional digital channel on the LANDSAT data tape. A soil code number is filled-in for each pixel so that a soil type can numerically be associated to each four element multispectral vector in the LANDSAT data. Figure 2.2-5 contains a gray scale image of the digital soil channel generated from the map for the area of the frame. Different gray levels correspond to different soil types although an exact correspondence cannot be illustrated in the image.

Land Use Map Registration

The second variable judged important in relating spectral strata to the physical environment was general land use. A land use patterns map consisting of twelve land use units transcribed onto a 1969 U.S.G.S. Base Map at a scale of 1:1,000,000 for the State of Kansas was obtained and processed.[6] A Lambert conformal conic projection is used for the map.

This map was similarly processed by mounting it on a coordinate digitizer and manually tracing out the edges of each land use unit and punching the coordinates on cards. These cards were then processed by a boundary registration system and a digital channel generated with land use codes for each pixel in the LANDSAT frame. The land use map is reproduced in Figure 2.2-6 and as can be seen it is much more complex than the soil map. There were approximately 300 arcs for the soil map compared to over 1000 for the land use map. Processing this large number of arcs caused problems in the software and resulted in extensive delay in completing the registration.

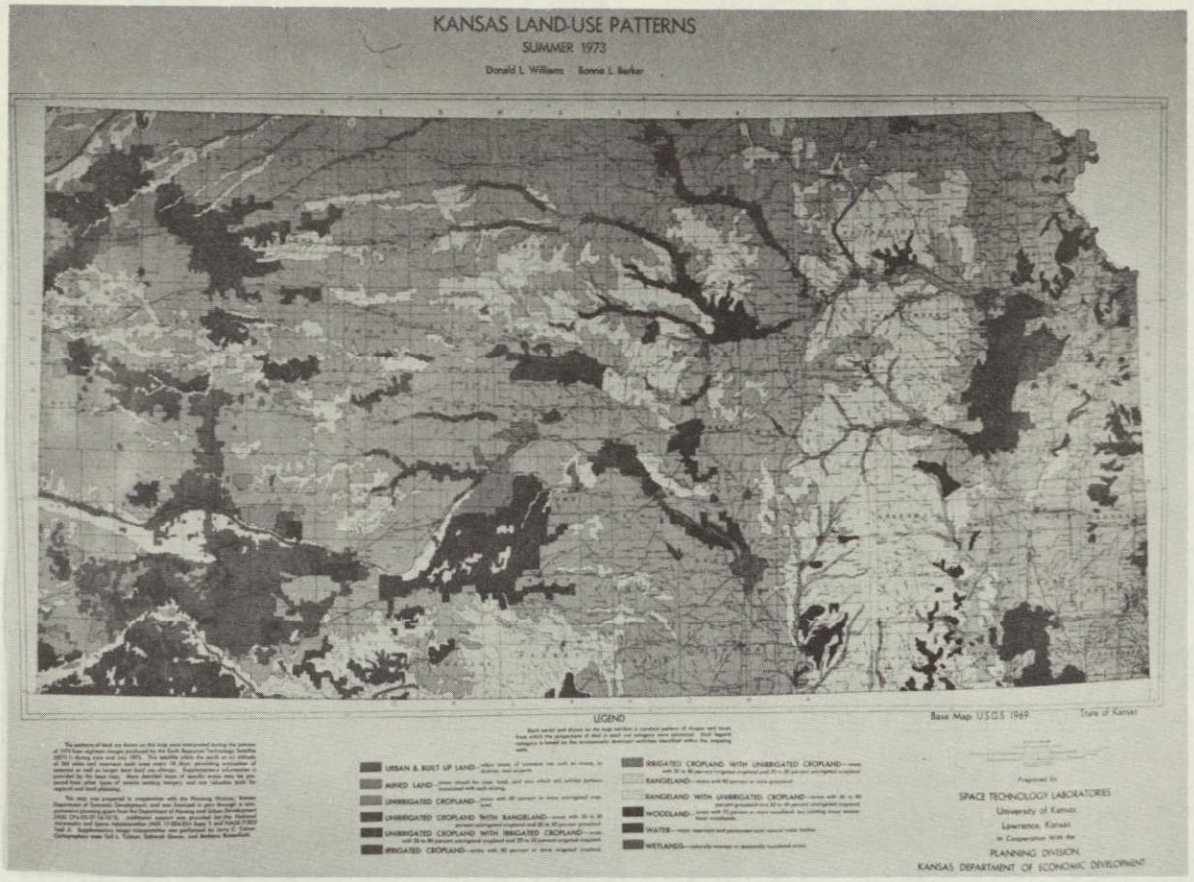


Figure 2.2-6. Source Map for Land Use Pattern Data.

Figure 2.2-5. Gray Scale Map of Soil Association Information for Scene 1689-16382, Acquired June 12, 1974.

Precipitation Data Registration

The precipitation data was available on only a county basis; thus the task was to digitize the county boundaries for Kansas and then insert the appropriate number in each county area. County boundaries were digitized from the soil map used previously and the digital county boundaries were combined with the county values to produce two additional data channels with a temperature or precipitation value for each pixel.

The precipitation data was obtained from data tapes supplied by the National Oceanic and Atmospheric Administration. The precipitation data were summarized on a monthly basis for each of the counties in Kansas. For the initial registrations the values used for each county were the total precipitation for the periods October 1, 1973 - June 30, 1974 and March 1 - June 30, 1974. Additional intervals of accumulation will be defined in CY77 to further explore the correlation of spectral properties with temperature as well as precipitation.

Outlook on the Use of Spectral Stratification

Large scale surveys using satellite-acquired multispectral data require classifications to be made over areas at least the size of individual LANDSAT scenes. The diversity of landscape patterns found over many areas of this size indicates that a logical first step in the analysis and classification of LANDSAT data is to stratify or divide the scene into units which are internally similar. Such a stratification will be helpful in constructing sampling frames which minimize the variance among sample units and in determining the boundaries of areas over which training statistics can be satisfactorily extended.

Stratification for sampling purposes can be based on static factors whose boundaries are either static or change only very slowly. For classification, however, the stratification should be based on the LANDSAT spectral data and will include the effects of dynamic as well as static factors.

The use of computer-implemented clustering procedure for dynamic stratification has been developed and tested over several LANDSAT scenes of Kansas. Initial results indicate that the technique can be used to determine the similarity of sample units and that the strata produced agree with major physical factors. The use of such a procedure should enable scenes to be more efficiently and objectively stratified than would be possible using manual methods.

We recommend that stratification be considered a prerequisite of signature extension or signature adjustment algorithms such as the multiplicative and additive signature correction (MASC) technique described by Henderson.[7] Our observation of results from such algorithms is that the results are highly variable and are data dependent. This shortcoming may be largely overcome by applying such signature adjustment algorithms only within a stratum, thus taking advantage of the knowledge gained from spectral stratification.

REFERENCES

1. Cochran, W.G. 1963.
Stratified Random Sampling.
In Sampling Techniques.
John Wiley and Sons, Inc., New York, pp. 87-113.

2. Wacker, A.G. and D.A. Landgrebe, 1972.
The Minimum Distance Approach to Classification.
Information Note 100771, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

3. Davis, B.J. and P.H. Swain. 1974.
An Automated Repeatable Data Analysis Procedure for Remote Sensing Applications.
Lars Information Note 041574.
Proceedings, Ninth International Symposium on Remote Sensing of Environment, April 15-19, 1974.
Ann Arbor, Michigan, pp. 771-774.

4. Kettig, R.L. and D.A. Landgrebe. 1975.
Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects.
IEEE Transactions of Geoscience Electronics, GE-14: 19-26.

5. Bidwell, O.W., and C.W. McBee. 1973.
Soils of Kansas.
Kansas Agricultural Experiment Station, Department of Agronomy Contribution No. 1359.

6. Williams, D.L., and B.L. Barker. 1974.
Kansas Land Use Patterns.
Space Technology Laboratories, University of Kansas, Lawrence, Kansas.

7. Henderson, R.G. 1976.
Signature Extension Using the MASC Algorithm.
IEEE Transactions of Geoscience Electronics, GE-14: 34-37.

ACKNOWLEDGMENTS

Marvin Bauer has been the leader of the spectral strata research and Barbara Davis has served as the project manager. The contributions of Tina Cary and Philip Swain in defining the clustering procedures and of William Pfaff in programming them are gratefully acknowledged. The classifications were performed by Paula Pickett and Ronald Boyd and Paul Anuta, with Nim-Yau Chu, carried-out the data registration tasks.

2.3 Field Measurements for Remote Sensing of Wheat

INTRODUCTION

Major advancements have been made in recent years in the capability to acquire, process, and interpret remotely sensed multispectral measurements of the energy reflected and emitted from crops, soils, and other earth surface features. With the initiation of experiments such as the Large Area Crop Inventory Experiment (LACIE) the technology is moving rapidly toward operational applications. There is, however, a continuing need for quantitative studies of the multispectral characteristics of crops and soils and the agronomic and meteorological factors influencing them if further advancements in the technology are to be made. In the past many such studies were made in the laboratory because of a lack of instrumentation suitable for field studies. However, the applicability of such studies is generally limited. The development of sensor systems capable of collecting high quality spectral measurements under field conditions now makes it possible to pursue investigations which would not have been possible a few years ago.

A major effort was initiated in the fall of 1974 by NASA/JSC with the cooperation of USDA to acquire fully annotated and calibrated multitemporal sets of spectral measurements and supporting agronomic and meteorological data. Spectral, agronomic, and meteorological measurements are being made on three LACIE test sites in Kansas, South Dakota, and North Dakota. The remote sensing measurements include data acquired by two truck-mounted spectrometers, a helicopter-borne spectrometer, an airborne multispectral scanner, and the Landsat-1 and -2 multispectral scanners. These data are supplimented by an extensive set of agronomic and meteorological data acquired during each remote sensing data collection mission. Together these data form one of the most complete and best documented data sets ever acquired for remote sensing research. Thus they are well-suited to serve as a data base for research defining future remote sensing systems and to quantitatively determine the spectral-temporal characteristics of spring and winter wheat crops.

At the beginning of the project, Purdue/LARS was requested to provide the technical leadership and coordination for the project as well as assume a major responsibility for the acquisition, processing, archiving, and analysis of the data. This report summarizes the activities of LARS in carrying-out these functions during the past year. The final sections of the report include summaries of analytical results of three investigations being pursued with field measurements data and recommendations for the future direction of the field measurements project.

DATA ACQUISITION AND MEASUREMENT TECHNIQUES DEVELOPMENT

Data acquisition activities were focused on the spring wheat crop in Williams County, North Dakota. The principal measurement sites were located at the North Dakota State University Agricultural Experiment Station near Williston, North Dakota. Spectral bidirectional reflectance factor, radiant temperature, agronomic crop parameters, meteorological conditions and total irradiance were measured for 60 plots arranged as shown in Figure 2.3-1. The treatments were chosen to represent the major agronomic factors affecting the growth, development, and yield of spring wheat. A summary of these measurements is shown in Table 2.3-1. Three of the plots were also used for measuring the thermal properties of wheat canopies; the results in the section on data analysis and results.

In addition, a large commercial wheat field was used for canopy modelling measurements. At this site, spectral bidirectional reflectance factor and radiant temperature were measured at view angles distributed over the complete hemisphere of observation for nearly fixed sun positions. Other data acquired for canopy modelling included photographs, agronomic data and ERTS-band reflectance factor measurements.

The following sections describe the instruments and procedures used in acquiring the data, the development of new instruments, and the development of improved field measurement techniques and procedures.

Radiometric and Reflectance Data Acquisition

The Purdue/LARS field data acquisition system (mobile tower, instrument van, and portable generator) was used to operate the Model 20C Spectroradiometer and precision radiation thermometer (PRT-5) above the plots on the experiment farm, the "modelling" field and the canvas calibration panels located at the helicopter site. This system also provided a maintenance and calibration facility for the instruments discussed below.

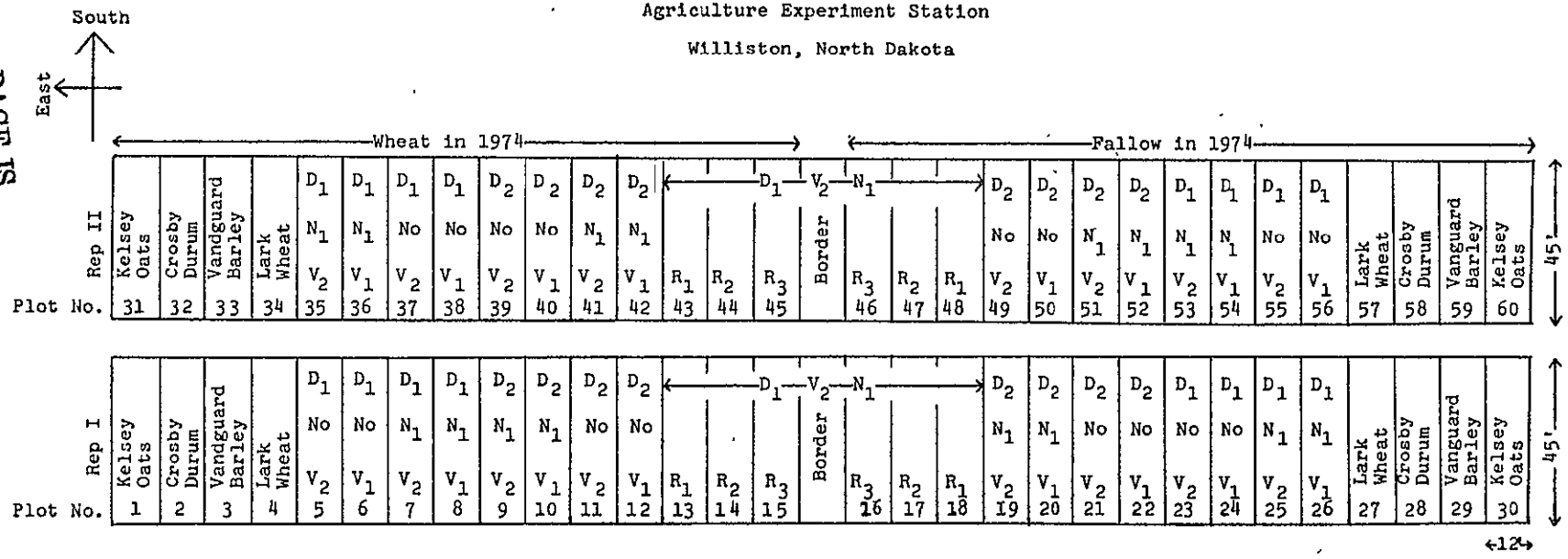
1. Calibration Procedures. A painted barium sulfate standard was used as the reflectance surface for spectral bidirectional reflectance measurements. For measurements in the thermal spectrum, two blackbodies were used to bracket the spectral radiance of the subject scenes. One of the blackbodies was used to calibrate the PRT-5 for temperatures from 15° C to 50° C.

2. Model 20C Field Spectroradiometer. The spectral bidirectional reflectance factor from 0.4 to 2.4 micrometers of the experimental plots and the canvas calibration panels was measured (15° field of view) with sun zenith angles less than 45° and view angle of 0° (straight down). For the modelling field, spectral bidirectional reflectance factor measurements were made over the hemisphere of observation at azimuth angles of 0, 45, 90, 135, 180, 225, 270, and 315 degrees with view angles

REMOTE SENSING -- WHEAT EXPERIMENTS

Agriculture Experiment Station

Williston, North Dakota



Treatment Descriptions

Soil Moisture	Planting Date	Variety	Nitrogen	Seeding Rate
M_1 Wheat in 1974	D_1 May 20	V_1 Olaf (Semi dwarf)	No None	R_1 30 lbs/acre
M_2 Fallow in 1974	D_2 May 30	V_2 Ellar	N_1 30 lbs/acre	R_2 60
				R_3 90

Figure 2.3-1. Experimental plots and treatments being studied by Purdue/LARS at Williston, North Dakota.

Table 2.3-1. Summary of data acquired by Purdue/LARS during 1975 at Williston, North Dakota.

Data	Wheat Growth Stage	Target Description Reflective Data	Thermal Data
June 1	Emergence		Plots 29, 30
5	Seedling	Plots 20-30, Cal panels	Plots 20-30
7	"	Plots 40-60, Fallow, Grass	Plots 55-60, Fallow
22	---	Cal panels	
24	Tillering	Modelling field-47 obs.	
July 9	Boot	Plots 31-41, Cal panels	
10	"	Plots 1-60, Grass, Winter Wheat, Fallow	Plots 52-60
11	"	Modelling field-57 obs.	
18	Heading	Cal panels and grass with ERL	
19	"	Plots 31-60, Fallow	
20	"	Modelling Field-29 obs.	
21	"	Plots 52-60, Grass	Plots 52-60
27	Headed	Plots 31-60, Cal panels	
28	Mature	Winter Wheat	
29	Headed	Plots 1-60, Fallow, Alfalfa	Plots 52-60
Aug. 12	Dough	Plots 1-60, Fallow, Alfalfa	Plots 52-60 (twice)
13	"	Modelling Field-129 obs.	
15	Ripening	Plots 13-30, Cal panels, Grass, Fallow and Alfalfa	
23	Mature	Plots 9-12, 20-22	

of 0, 15, 30, 45, and 60 degrees from normal. The reference for these observations was viewed at 0^{o}. These measurements were made over the entire day, when possible, providing data for sun zenith angles distributed over a wide range. The model 20C spectral reflectance factors were corrected for the reflectance of Barium Sulfate.

The thermal unit of the Model 20C was used to measure the spectral directional radiance in a 15^{o} field of view for selected experimental plots. Measurements were performed with a view angle of 0^{o} simultaneous with spectral reflectance factor and PRT-5 measurements. Thermal spectral measurements of these plots were completed within ten minutes under stable wind and incident light conditions.

During all measurements the Model 20C and PRT-5 were positioned at a height of seven meters above the canopy. The motor drive camera was aligned with the reflective unit to acquire photographs of each plot during the measurement. Additionally, oblique photographs of each plot were taken using a hand-held camera at a height of approximately two meters.

3. PRT-5 Precision Radiation Thermometer. This instrument was used to view the plots during spectral reflectance factor measurements. Data were recorded manually and entered in the data bank with the description of the measurement which is recorded for each reflectance spectra.

4. Model 100 ERTS-Band Radiometer. The spectral-band reflectance factor (in ERTS/Landsat bands) of four plots within the modelling field was measured using a tripod based mounting beam to position the radiometer at a height of 2 meters. The 15^{o} field of view was used with a view angle of 0^{o}. Data were acquired on June 24, July 11, 20, and 30, and August 13. The reflectance factor relative to a painted Barium Sulfate panel was measured several times during the day for various sun angles. The directional-hemispherical transmittance of green, yellow, and brown leaves was measured using the transmittance attachment developed at LARS during the previous year.

Photographic records of the canopy included color slides taken at angles of 0, 10, 20, 30, 40, 50, and 60 degrees perpendicular and parallel to the row direction and silhouette photos of single rows and single plants.

5. Total Incidence Pyranometer. An Eppley Model 8-48 pyranometer was stationed near the plots on the experiment station. Total irradiance versus time was recorded on a strip chart for each day on which spectral data were recorded.

Meteorological Data Acquisition

Meteorological data were measured and recorded at the experiment station during each day on which spectroradiometric data were acquired. The measurements were acquired using the meteorological data acquisition system designed and implemented the previous year. Measurements of barometric pressure, temperature, relative humidity, and wind speed and direction are recorded on strip charts using instruments manufactured by Weather Measure, Inc.

Measurements of temperatures at various positions in and above wheat canopies were accomplished using a system developed at LARS in FY75 and 76. Five thermilinear probes for measuring air and contact temperatures were deployed from a single post and up to nine posts were multiplexed to a temperature to voltage converter for measurement using a digital voltmeter.

Agronomic Data Acquisition

Agronomic observations and measurements describing the condition of each of the plots were made at the time of the spectral data acquisition. These data include: maturity, height, percent ground cover, leaf area index, biomass of leaves, stems, and heads, soil moisture and condition, and grain yield. These data are supplemented by vertical and oblique photographs of each plot.

Instrument and Procedures Development

A number of developmental tasks were required in support of the data acquisition activities described above. These tasks included:

1. Documentation of the directional-hemispherical transmittance attachment to the Model 100 ERTS-Band Radiometer was completed. [LARS Information Note 052075] Extensive tests were performed on the transmittance measuring system to determine the comparability with Beckman DK-2 transmittance measurements. Source and system transfer characteristics of both instruments were digitized directional-hemispherical transmittance of typical leaf samples was measured by both instruments, and the results were compared. Documentation of this evaluation, in process, indicated that the instruments compare favorably.

2. Development of a directional-hemispherical reflectance attachment for the Model 100 ERTS-Band Radiometer was begun. Field tests of the attachment in May of 1976 indicated that further development is required. Modifications were made and further tests will be performed in June of 1976.

3. Field performance evaluations were performed for the truck spectrometers involved in the Field Measurements Project. At the request of NASA/ERL, tests were performed on their Model 20D system to determine the cause of reflectance values which were inconsistent with previous performance. The causes were established and a procedure to determine the instrument behavior over the summer was developed. This enabled ERL personnel to correct the data.

In cooperation with NASA and Lockheed personnel, the operational procedures for the Field Signature Acquisition System were examined and it was determined that the number of observations per plot could be significantly reduced without reducing the quality of the data. This and other modifications of operational procedures enabled the operation crew to acquire high quality data at a rate consistent with the requirements of the experiment.

Field procedures used routinely to check the performance of the LARS Model 20C were demonstrated to the operation crews using their instruments. Frequent, on site, tests especially field of view tests, were shown to be very important to the acquisition of quality data. As well, the measurement of the spectral bidirectional reflectance factor of the gray panels by all instruments proved to be of great value in the production of comparable data.

4. Techniques developed at NASA Goddard and adapted at Purdue for the application of Barium Sulfate in a durable, highly reflecting coating for reference surfaces for field instrumentation were used to produce reference panels for spectral reflectance factor calibrations. Four 1.3 meter square panels were painted for use by the truck spectrometers, two 0.6 meter square panels were prepared for the modelling measurements, two 0.3 meter diameter panels were also pointed for the JSC/FSAS system.

5. Construction of an eight channel printing data logger was completed and conditioning cards were prepared to acquire

- five channels of temperature data from 0^{o} to 45^{o} C using Thermilinear probes
- four channels of Model 100 ERTS-Band Radiometer data and one channel of PRT-5 data
- eight channels of unconditioned data from 0 to 2.0 or 0 to 20 volts.

Major features of the data logger include: two digits of sequential data coding, one digit indicating the channel being printed, and $3\frac{1}{2}$ digits of measured data. The instrument will measure and print the data in a selected channel or scan and print sequentially up to and including a selected channel at a rate of about three channels per second. More complete documentation of the system is in progress.

6. Documentation was nearly completed for the bidirectional reflectance factor reflectometer designed to use the Exotech Model 20C and a high intensity source to make indoor measurements of large area samples. The device facilitates the comparison of pressed Barium Sulfate with the painted Barium Sulfate panels and has provided the basis for three investigations of the reflective characteristics of soils. The document discusses the bidirectional nature of reflectance in accepted definitions and symbols. Results for typical sample surfaces--paints, soil, and cloth--are presented in several formats which are convenient for physical interpretation.

DATA PROCESSING AND REFORMATTING

After the data have been acquired in the field they must be prepared for analysis. It is important that the data be calibrated consistently and made available in comparable formats in order to make meaningful comparisons among data acquired by the several sensors at different times and locations. A second key step in the processing of all the data is to correlate the agronomic, meteorological, and photographic data with the spectral data and as fully as possible integrate the various data types. To accomplish this the agronomic and meteorological data are keypunched and entered on the digital tapes along with the spectrometer data.

This section summarizes the data processing and reformatting steps performed by LARS and reports on the current status of the data.

Procedures

The processing functions performed not only by LARS, but also by NASA/JSC and NASA/ERL have been previously described in detail in the Field Measurements Project Status Report prepared in January 1976 by LARS. Thus, they will only be briefly summarized in this report. Figures 2.3-2 and 2.3-3 illustrate the major steps in processing the spectrometer data. The spectrometer data is calibrated to reflectance factor using known Barium Sulfate standards and the data for each sensor is put in the same wavelength bands to facilitate comparisons among data from the several spectrometer systems being used in the project. The wavelength band intervals are 0.02, 0.05, and 0.10μm for the regions 0.4-1.4, 1.4-2.4, and 2.7-14.0μm, respectively.

The multispectral scanner data processing is primarily reformatting from Universal to LARSYS format. The Landsat data are being multitemporally registered to facilitate their analysis.

Status

The status of preparation of 1974-75 mission data is summarized in Table 2.3-2. Processing of the 1974-75 data has been delayed while software for handling the newer types of data and putting data into the field measurements format have been prepared. Also, a number of unexpected problems in instrument performance occurred which have required additional processing to correct. Software systems are now complete and all processing of 1974-75 data will be completed by August 1976. Processing of 1975-76 data has already begun and is expected to proceed much more smoothly and quickly since the necessary software and procedures have been developed and tested.

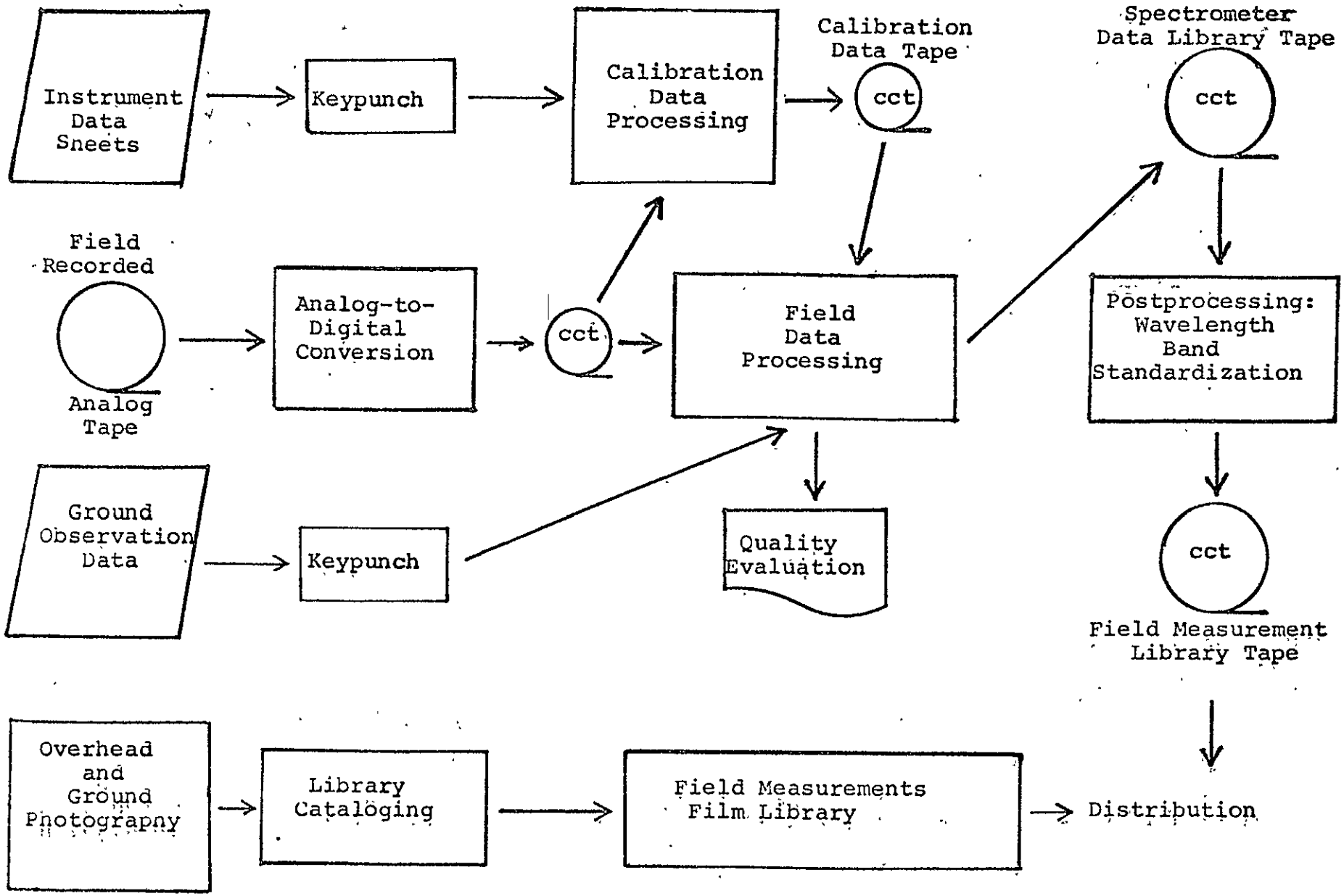


Figure 2.3-2. LARS Field Spectroradiometer Data Processing

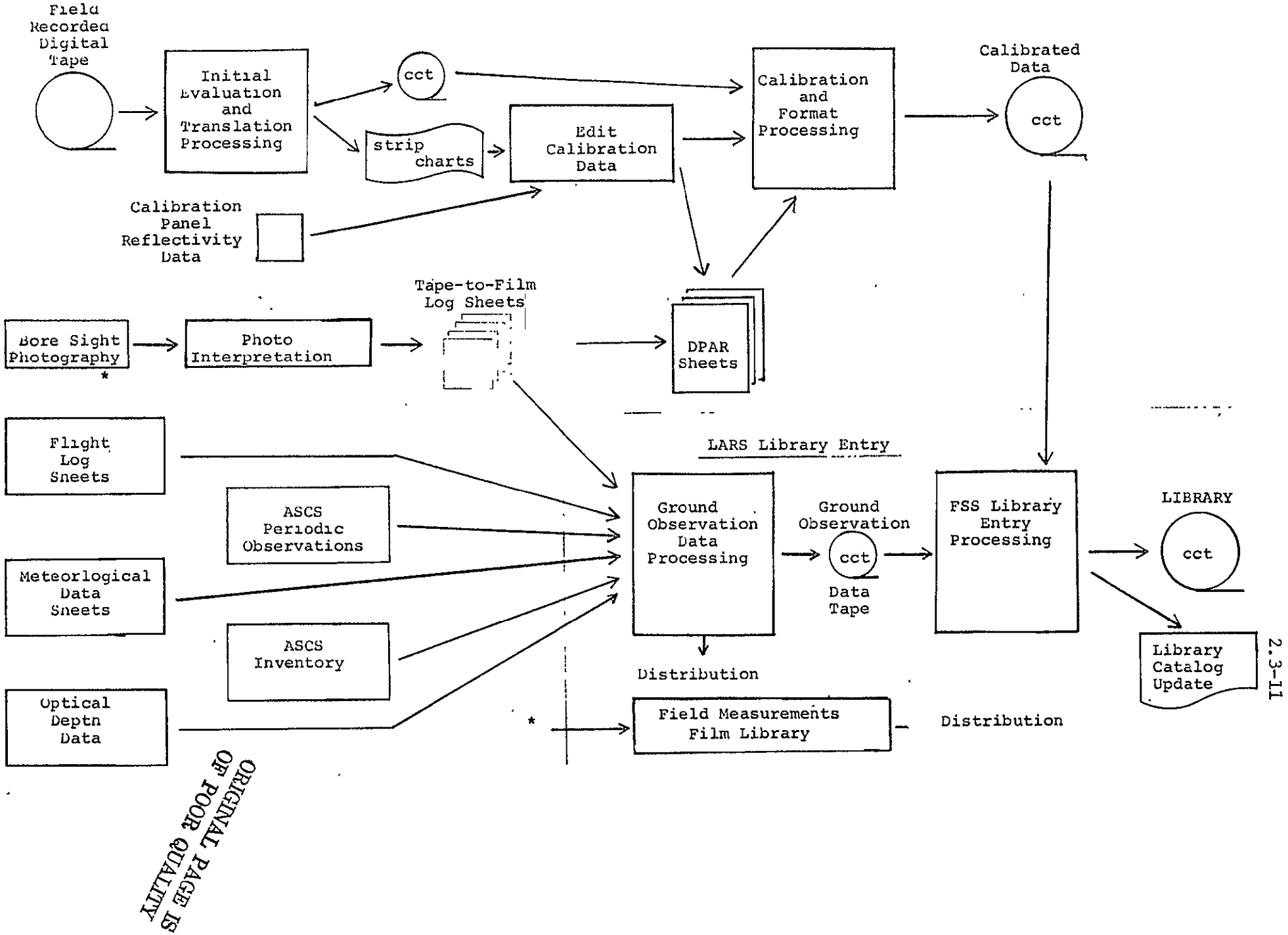


Figure 2.3-3. S-191 Data Processing

Table 2.3-2. Summary of 1974-75 Field Measurements Data Preparation Status.

Landsat MSS Data, CCT's and Imagery
30 frames received and in library

MSS and MMS Aircraft Scanner Data
16 missions received at LARS
6 missions reformatted
10 missions available upon request

FSS (S-191) Data
20 missions received at LARS
3 missions final processing complete and in library
17 missions partially processed

20C Spectroradiometer Data
6 missions final processing complete and in library
3 missions partially processed

20D Spectroradiometer Data
2 missions final processing complete and in library
13 missions partially processed

FSAS Interferometer Data
3 missions received, processing complete, and in library

ASCS Inventory and Periodic Observation Data
23 missions received and keypunched

Meteorological Data
18 missions received and keypunched

DATA QUALITY VERIFICATION AND EVALUATION

An important part of the field measurements project is the acquisition of high quality, calibrated spectral measurements. To a large degree, this depends on having a timely and quantitative method available for determining data characteristics. This information can be used (1) for identifying sensor deficiencies which can be corrected and (2) by data analysts in selecting data for analysis and in interpreting analysis results. The importance of quantitative data verification has become increasingly apparent to LARS as the project has progressed and this section summarizes our initial efforts to strengthen this part of the project.

In April 1976 a document describing our technical recommendations for data quality evaluation and verification for the spectrometer and aircraft scanner systems was prepared. The key points of those recommendations are presented along with our current evaluation of data quality.

Data Quality Evaluations

The uncertainity of measurement for each of the spectrometer systems has been determined. For reflective data, the values are 7, 10-15, and 8-29 percent for the LARS, ERL, and S-191 spectrometer systems, respectively. It has been demonstrated that the LARS system produces repeatable data from mission to mission. The high value for the ERL data is due to misalignment of the first minor, but the system was stable and it has been determined that the solar port data can be used to compute reflectances based on calibrations before and after the occurrence of the problem. The problem has been corrected in the data processing and the final data is expected to be similar to the LARS data.

The FSS (S-191) data has exhibited an unusually high degree of uncertainity for certain missions which are attributed to collection of data during partially cloudy periods, missing the calibration panel, or instrument malfunctions. The latter is considered least likely and steps have been taken to reduce the re-occurrence of the first two problems in 1976 missions.

The 24-channel MSS data has been plagued by data quality problems including banding, bit errors, saturation, and inoperative bands. By averaging pixels, i.e., degrading the spatial resolution, and other preprocessing stops, the data quality has been substantially improved and high classification performances have been achieved with the data. Insufficient amounts of MSS data have been received to evaluate its quality.

Data Quality Determination

Recommendations for determining data quality have been prepared and submitted to NASA/JSC. Some of the initial steps would be carried-out by JSC, with the final quantitative steps performed by LARS. The key steps are summarized here for the aircraft scanner, helicopter spectrometer, and truck-mounted spectrometers.

1. Aircraft Multispectral Scanner

The first step in collecting high quality data in the field is determined acceptable limits of illumination. We have recommended the use of a recording total incidence pyranometer to determine quantitatively and at a glance whether illumination meets the minimum requirements.

Histograms of data can be used to determine detector operation, bit dropping or favortism, dynamic range, sensitivity, and saturation. Several quantitative measures based on analysis of histograms of each data channel have been defined. These include: (1) serial correlation coefficient to determine low order bit errors, (2) peak relative change in frequency to determine high order bit errors, and (3) percentage of saturated data points to detect improper range and/or offset settings. In addition, a comparison of $NE\Delta\rho$ and $NE\Delta T$ with sensor performance specifications as a measure of system noise, and determination of the number of line synchronization errors is recommended. Threshold levels denoting unacceptable performance have been defined for each of these measures. Further tests to determine the correspondence between the data quality indicators and classification performance are required and should be carried-out during the coming year.

2. FSS (S-191) Data

Major factors contributing to the uncertainity of helicopter spectrometer data are, in order of importance, clouds, calibration procedures, atmospheric conditions, instrument performance and, data processing system performance. The requirement for sensor performance evaluation has long been recognized and quick-look and other, more extensive, instrument tests are routine. To establish a measure of the uncertainty and causes of uncertainty throughout the system, the following steps have been recommended:

(1) A record of the total irradiance at the helicopter calibration site as a function of time to determine the illumination conditions.

(2) A calibration uniformity test. After a cosine correction for sun angle the calibration spectra should be nearly identical.

(3) Histogram tests for the processed data to indicate proper A-D conversion. These tests would be similar to those described for the scanner data.

(4) A system performance test utilizing measurements of a series of gray-scale calibration panels.

3. Truck Spectrometer Data (Exotech 20C, Exotech 20D, FSAS)

Major factors contributing to the uncertainty of truck spectrometer data are, in order of importance, calibration and operation procedures, calibration frequency, instrument performance, atmospheric conditions, and data processing system performance. In the case of the FSAS interferometer system it is diffcult to separate the data processing from the instrument; but, with software changes and hardware refurbishing and improvements, operational and calibrational procedures should become the limiting factors on the performance of this system. To establish a measure of the uncertainty and causes of uncertainty throughout these systems, the following steps have been recommended:

(1) Operational and calibration procedures which include (a) frequent checks of alignment of the field of view and (b) calibration using large Barium Sulphate panels.

(2) A record of total irradiance at the site as a function of time.

(3) A calibration uniformity test similar to the one described above for the S-191 data.

(4) Histogram tests for verifying analog to digital conversion similar to the ones described for the scanner data.

(5) A system performance test including measurements of gray-scale calibration panels. Spectra would be examined for continuity, form, and magnitude.

DATA LIBRARY

The field measurements data library contains a broad spectrum of information collected for the three agricultural test sites in Finney County, Kansas, Williams County, North Dakota, and Hand County, South Dakota. Data is collected at the agriculture experiment stations and the LACIE intensive test sites in Finney County and Williams County; but, only over the LACIE intensive test site in Hand County. For the Finney County and Williams County sites, canopy modelling data is also collected over a commercial wheat field.

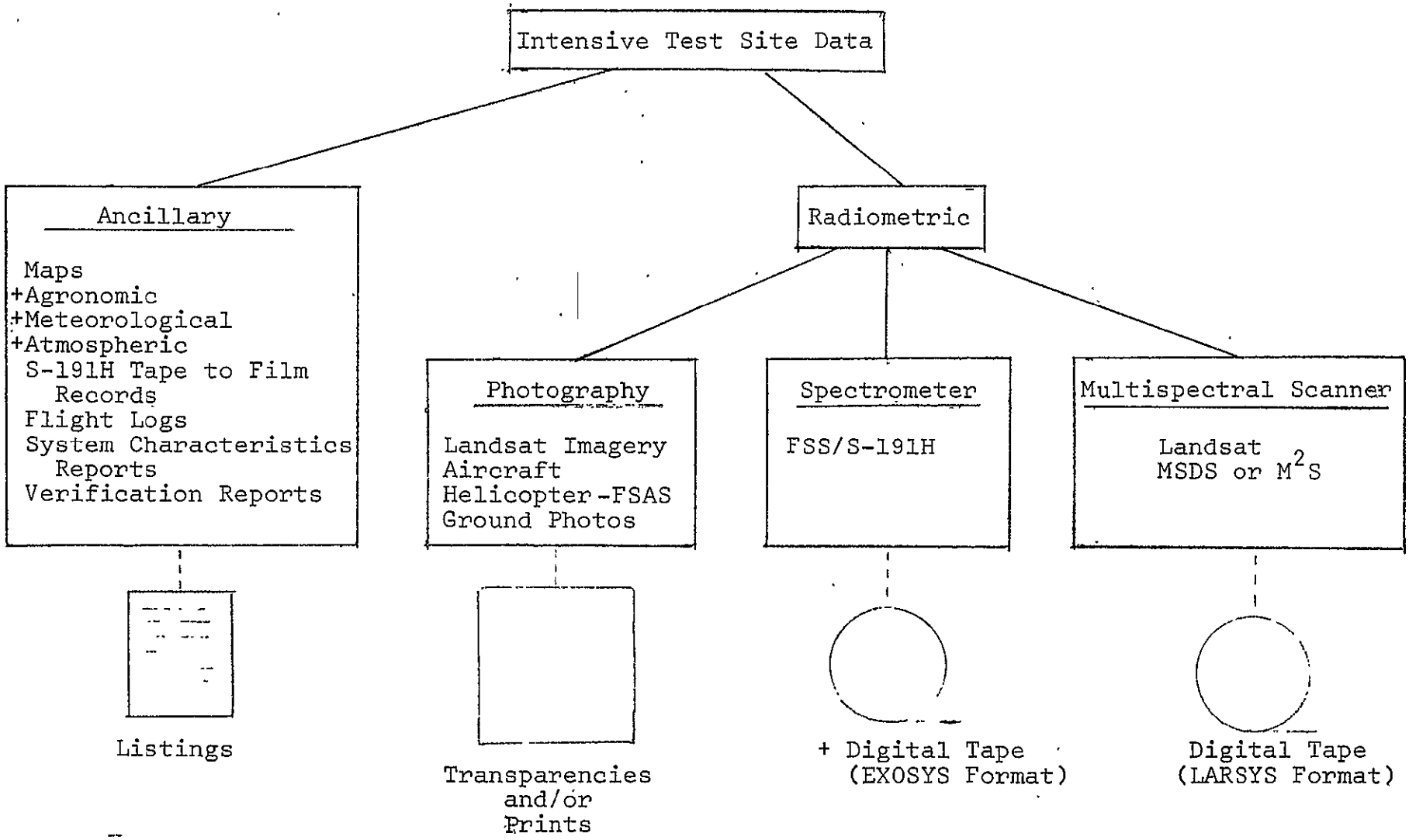
LARS is the central facility for storage and dissemination of field measurements data being collected over the above three test sites. The organization of the library is shown in Figures 2.3-4, 5, and 6.

A computerized catalog listing of all data in the library is currently being prepared and should be available by August 1976. The listing will contain key items of information for each data run. A run is an individual observation or scan of spectrometer data, a flight-line at a given altitude of aircraft scanner data, or for Landsat data the entire test site. The items describing each run include: mission date, time, location, crop, sensor, and type of data (reflective and/or thermal), plus the run number uniquely identifying each item of data.

A summary of the spectral data currently in the library is shown in Table 2.3-3. The spectral data for each mission is supplemented by photographic, agronomic, and meteorological data acquired during each mission. The types of data acquired have been described previously in the project plan and the January 1976 status report.

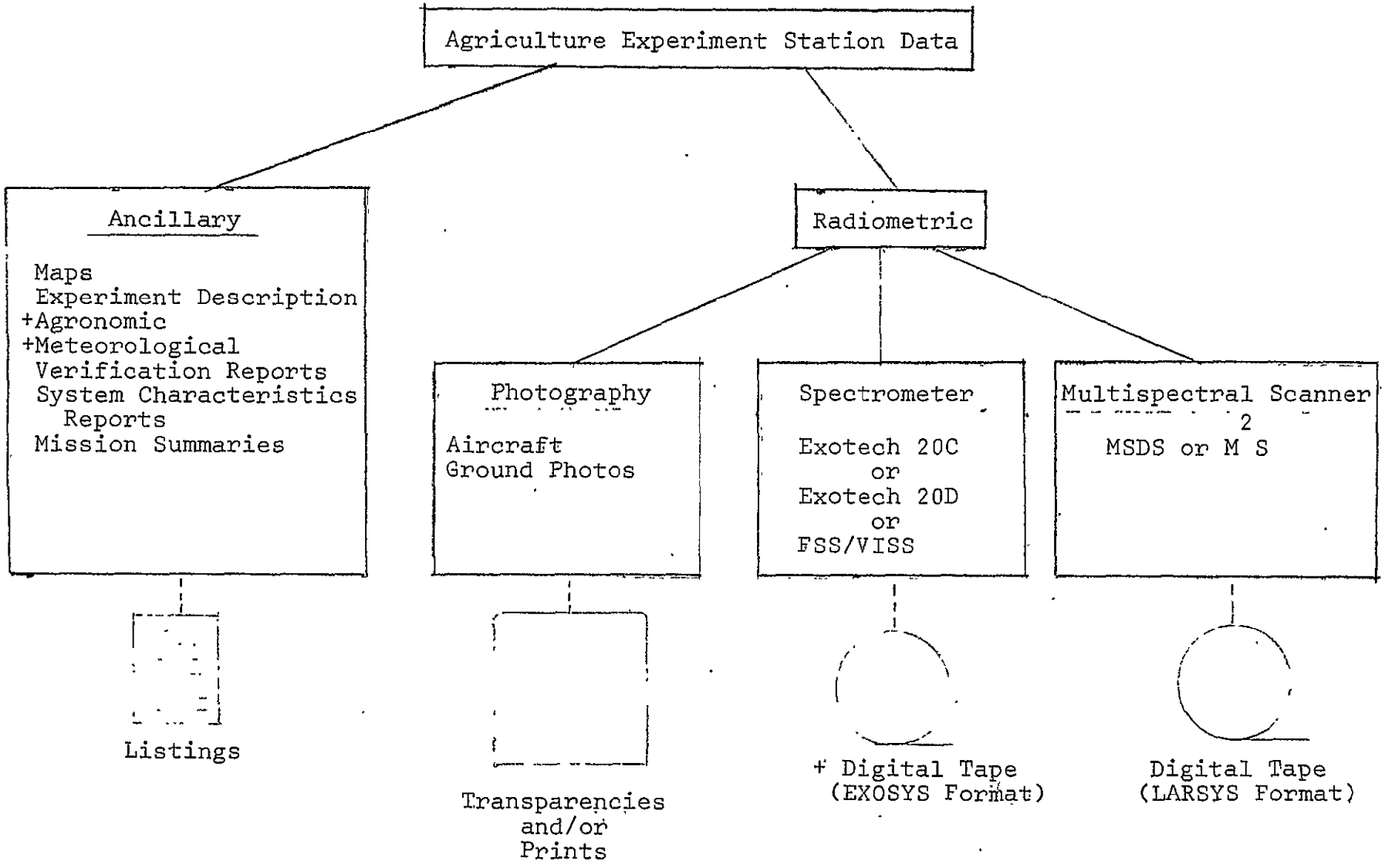
During the past year the following data have been distributed to researchers:

- Selected bands of 24-channel MSS data for four missions to ERL, GISS, and GSFC. (See section 2.4 of this report).
- S-191 data for eight missions to Texas A&M and two missions to ERIM.
- Exotech 20C (LARS) data for two missions to Texas A&M and ERIM.



\+ Ancillary data that are also stored as identification information on digital tape

Figure 2.3-4 List of types of data collected over the Intensive Test Sites and their final formats.



Ancillary data that are also stored as identification information on digital tape

Figure 2.3-5 List of types of data collected over the Agriculture Experiment Stations and their final formats.

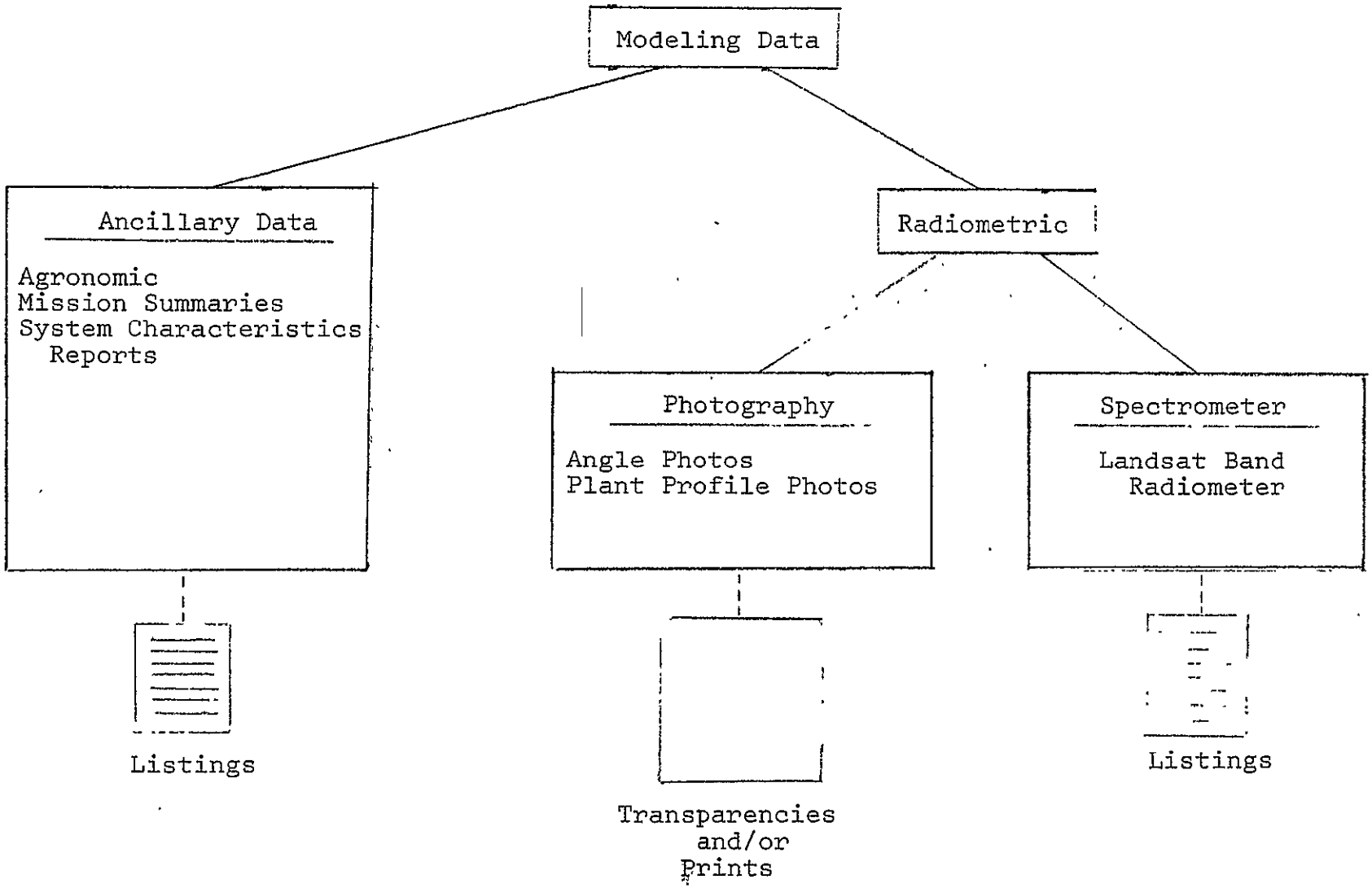


Figure 2.3-6 List of types of data collected over the modelling fields and their final formats.

Table 2.3-3. Summary of remote sensing data in the field measurements data library. A full set of agronomic and meteorological data is also in the library for each mission.

Mission	Wheat Growth Stage	Sensor Type					
		Landsat 1	Landsat 2	A/C MSS	S-191	Truck	Modelling Data
Finney Co., Kansas							
Oct. 17-20, 1974	Seedling	X				X	
Nov. 4-7, 1974	Tillering				X	X	
Nov. 23-25, 1974	Tillering	X				X	
March 19-22, 1975	Tillering				X	X	X
April 6-9, 1975	Jointing			X*	X	X	
April 24-27, 1975	Jointing		X			X	X
May 13-16, 1975	Boot			X	X	X	
May 21-24, 1975	Heading	X		X	X	X	X
May 30-June 2, 1975	Milk			X	X	X	
June 8-11, 1975	Dough			X	X	X	
June 17-20, 1975	Ripening		X	X	X	X	X
June 25-28, 1975	Mature	X		X	X	X	
July 5-8, 1975	Post Harvest		X	X	X	X	
Sept. 14-17, 1975	Pre-emergence		X		X		
Oct. 2-5, 1975	Seedling		X		X	X	
Oct. 20-23, 1975	Seedling		X		X	X	
Nov. 11-12, 1975	Tillering			X	X		
March 13-19, 1976	Tillering		X	0	X		
March 30-April 2, 1976	Jointing		X		X	X	
April 18-21, 1976	Jointing			0	X	X	
Williams Co., North Dakota							
May 25-28, 1975	Emergence	X		X		X	
June 3-7, 1975	Seedling		X	X	X	X	
June 12-15, 1975	Seedling	X		X			
June 21-24, 1975	Tillering		X	X	X	X	X
June 30-July 3, 1975	Jointing			X		X	
July 9-12, 1975	Boot		X	X	X	X	X
July 18-21, 1975	Heading			0	X	X	X
July 27-30, 1975	Headed		X		X	X	X
Aug. 5-8, 1975	Milk-dough			0	X	X	
Aug. 14-17, 1975	Ripening		X	X	X	X	X
Aug. 23-27, 1975	Mature			X	X		
Sept. 1-4, 1975	Post Harvest			0	X		
Hand Co., South Dakota							
Sept. 25, 1975	Pre-emergence			0			
Oct. 15, 1975	Emergence				X		
Oct. 30, 1975	Seedling				X		
Nov. 5, 1975	Tillering			X			

* X=MSS, O=M^2S

TECHNICAL COORDINATION

A substantial amount of effort has been devoted to technical communication in order to coordinate all phases of the project. It would be virtually impossible to site all of the many exchanges of telephone calls and letters which have gone between NASA/JSC and LARS in the process of establishing data collection schedules, defining measurements, instrumentation and procedures, calibrating and processing data, and verifying and evaluating data quality. These were supplemented by visits of several LARS staff members to JSC as well as to the three test sites. Meetings were also held at Purdue in September and at Texas A&M in March with data users from ERIM, and Colorado State and Texas A&M Universities. As part of our technical coordination role, LARS prepared an update of the project plan in August 1975 and an overall project status report in January 1976.

DATA ANALYSIS AND RESULTS

Analytical results have been obtained for three experiments conducted with field measurements data acquired at Williston, North Dakota during the summer of 1975. Brief summaries of the major results will be presented in this report and complete, detailed reports of each of the experiments will be prepared in 1976.

Thermal Modelling of Spring Wheat Canopies

The primary objectives of the thermal modelling research conducted during 1975 at Willston, North Dakota were threefold: (1) to measure and identify the spectral characteristics of the radiance of crop canopies from 2.7-14.0μm that are relevant to remote sensing applications, (2) to measure the daily transient temperature profiles in selected crop canopies and interpret these profiles in terms of agronomic and environmental parameters on a temporal basis, and (3) to relate the measured spectral radiance to canopy temperature profile.

Using the observations on selected spring wheat canopies primarily in the green and mature stages of development, the thermal modelling analysis has treated three major topics which are discussed.

1. Spectral Radiance Characteristics in the Thermal Infrared

The spectral radiance, L_λ, of the test canopy was measured by the Exotech 20C spectroradiometer in the thermal infrared spectral region, 2.7-14 μm, using observation and calibration procedures described in previous publications by the LARS/Measurements group. In addition to spectral radiance as a function of wavelength, the data set contains all the relevant thermal, agronomic and environmental parameters that are necessary to more fully characterize the canopy.

The spectral radiance of a typical mature crop canopy for the wavelength region 2.7 to 14.0μm is shown in Figure 2.3-7. Physical interpretation of the observations in this form is difficult as only a few gross features can be discerned such as the wavelength at which the spectral radiance peaks and atmospheric absorption bands. It should be mentioned that the observations are generated in two spectral segments since the atmospheric water absorption effects between 5.5 and 7.0μm are so large.

To identify important physical features of the spectral radiance spectra, the following analysis was performed. The temperature, T_R, at which the Planck function, $L_b(\lambda,T_R)$, describing the spectral radiance, L_λ, at 4.6μm for the 2-6μm region and at 11.2μm for the 6-14μm region was determined. These matching wavelengths were selected on the basis that atmospheric absorption, target anomalies (the silica restrahlen phenomena for example) or reflected solar irradiance effects should be

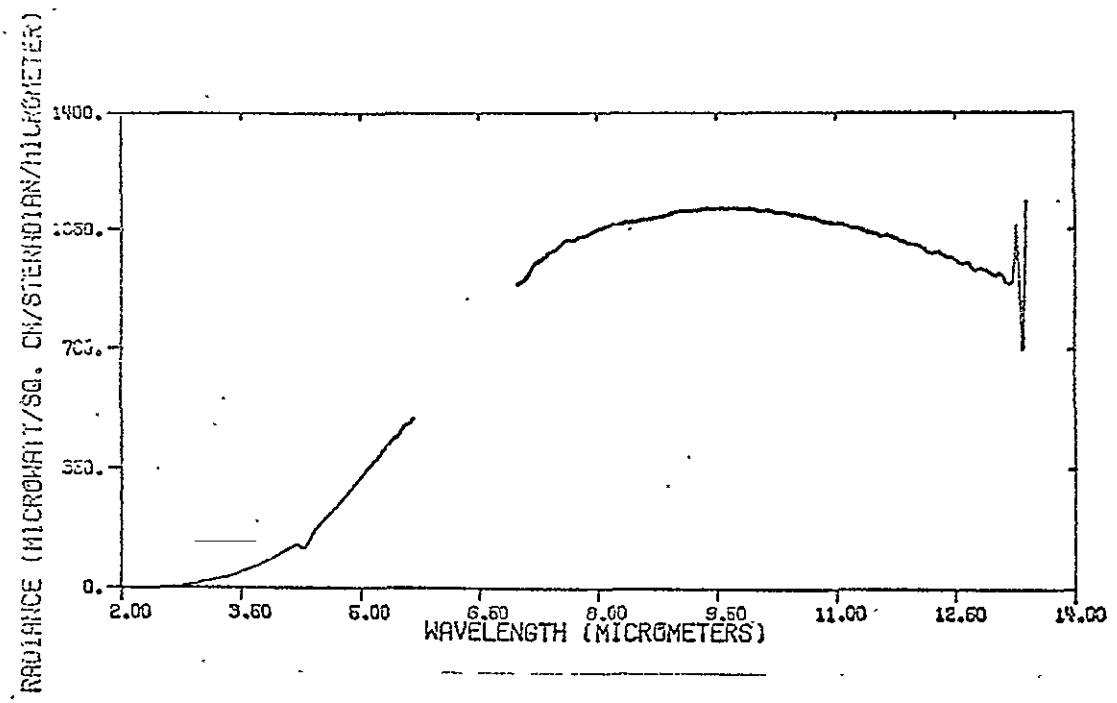


Figure 2.3-7. Observed Spectral Radiance of a Typical Mature Wheat Canopy (Plot 52, 1627GMT, July 1976).

minimal. The difference in the observed canopy spectral radiance, L_λ, and the Planck function, $L_b(_\lambda,T_R)$, was then shown as a deviation plot in Figures 2.3-8 and 2.3-9 where the deviation, DEV, is:

$$DEV = L_\lambda - L_b(\lambda T_R).$$

These deviation curves are amenable to physical interpretation. First, note the deviations are zero at the wavelength match points 4.6 and 10.5μm in Figures 2.3-8 and 2.3-9, respectively. In Figure 2.3-5, two two important effects can be seen; the positive deviation in the 3-4μm region indicates the presence of reflected solar irradiation while the negative deviations at 4.3μm and beyond 5.0μm can be explained by CO_2 and H_2O absorption bands. The deviation plot of Figure 2.3- for the spectral region 6-14μm is less simple to interpret. The positive deviations are likely to be atmospheric emission bands while the negative deviation is a restrahlen phenomena resulting from, in this case, the presence of silica in the soil which may constitute a portion of the target.

From more detailed consideration of these figures using procedures not reported in this abbreviated account, the following conclusions can be obtained from analysis of the deviation plots: (1) The spectral reflectance of the target can be estimated in the 3.5-4.0μm region; this has been done for bare soil (typically 13%) and for full green canopies (typically 3%). Analysis has been performed on selected canopies to show that these reflectance values correlate in a rational manner with leaf area index, percent ground cover, and biomass per unit area and (2) the observed spectral radiance in the 4.6-4.8μm spectral region results almost exclusively from the target emission. The consequences of the last two conclusions are very important and allow for the reliable determination of the spectral radiance temperature of the canopy. Further discussion on this will be presented in a later section.

2. Canopy Temperature Profile Measurements

The second objective of the past year's thermal modelling research activity was to measure temperature profiles in selected crop canopies and interpret these profiles in terms of agronomic and environmental parameters on a temporal basis. The purpose of these measurements was to learn in what manner temporal variations of leaf temperature, canopy air, and soil temperatures might be important to the remote sensing problem. The thermal radiation emitted by a canopy originates from surfaces (leaves, stems, and soil) that are at different temperatures and consequently the remotely sensed spectral radiance of the canopy will be dependent upon the temperature profile in a very complicated manner.

Three adjacent canopies (Plots 52, 53, and 54) were selected for profile temperature measurements in addition to detailed radiometric and agronomic observations. Using thermistor probes, Yellow Springs Instrument Co., Type 705, the temperatures with the canopy at the following locations were observed:

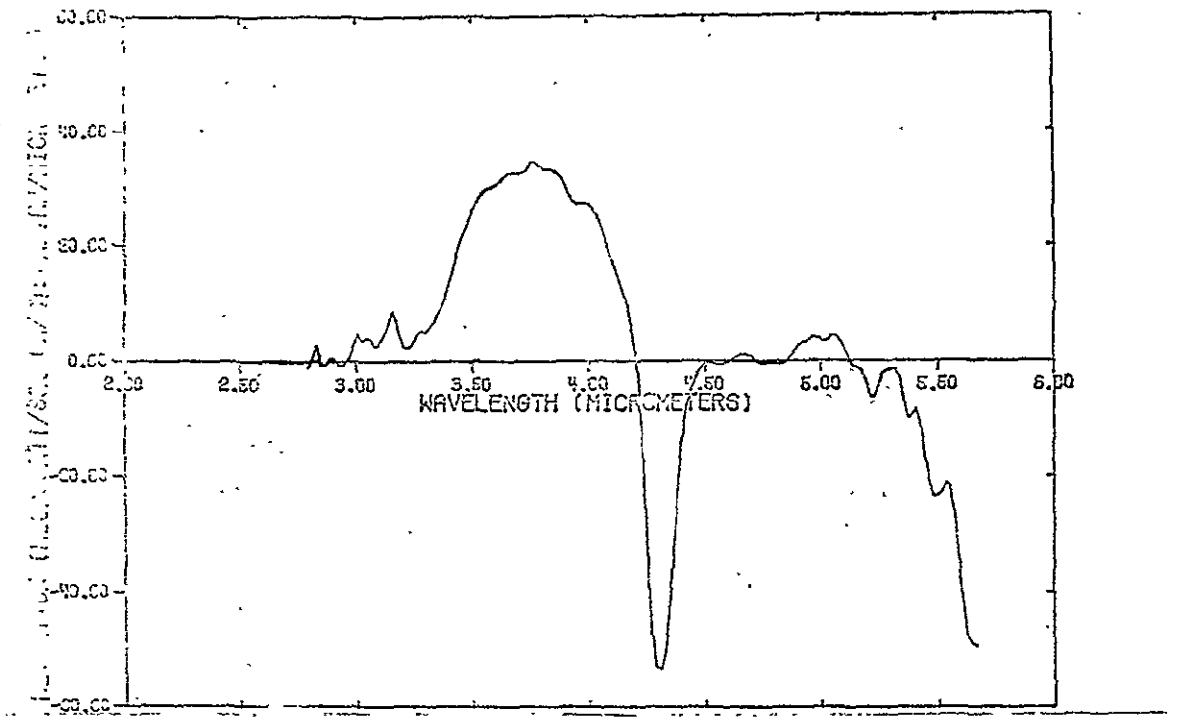


Figure 2.3-8. Deviation of Typical Canopy Radiance from Blackbody Radiance in the 2.8 to 5.6μm wavelength region

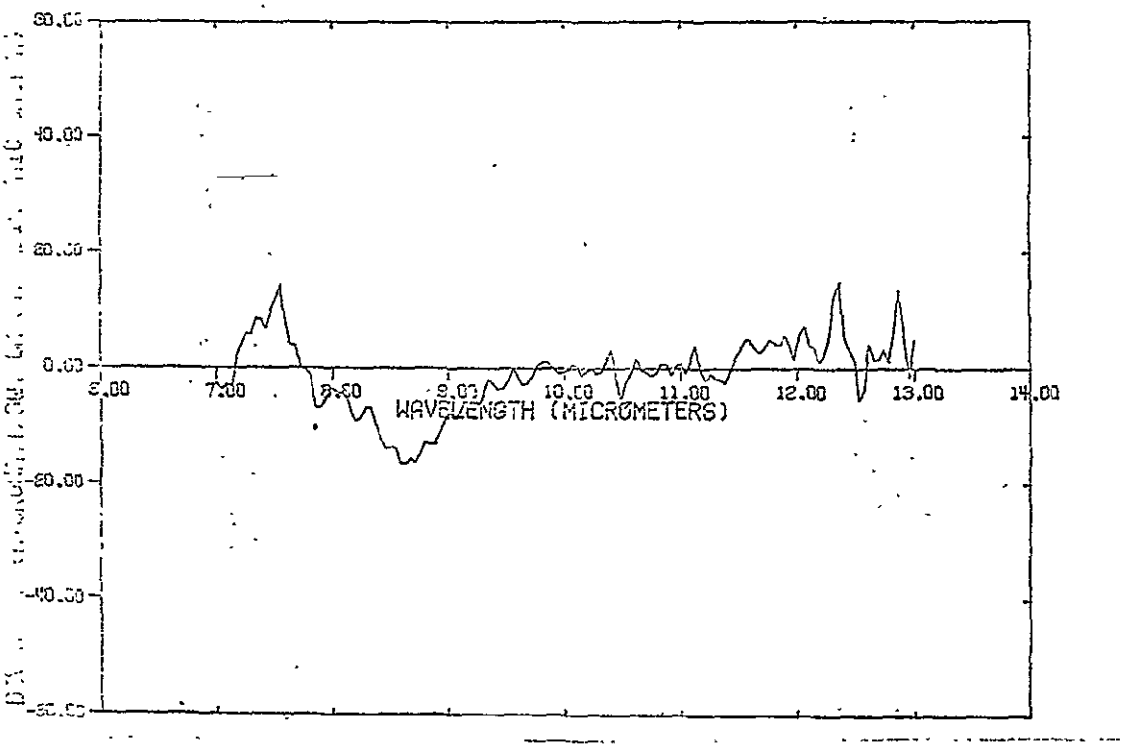


Figure 2.3-9. Deviation of Typical Canopy Radiance from Blackbody Radiance in the 7.0 to 13.3μm wavelength region.

Probe #	Location
1	Soil Surface, bead just covered with soil
2	Surface air temperature, approximately 5cm above the soil
3	First green leaf
4	Flag leaf
5	Air temperature, approximately 10cm, above the canopy

In some instances subsurface soil temperature at various depths were measured but not collected as a matter of routine. The probes described in the above table were mounted on a stake assembly which permitted easy placement in the canopy with minimum disturbance. In order to obtain a meaningful average three such stake assemblies were located on the circumference of a 3m diameter circle which bounded the radiometric observation target.

Figure 2.3-10 represents the temperature profile within a typical mature canopy (Plot 52, 27 July 1976) as a function of time. The probe positioning is as described in the above table; in addition two points at 1650GMT and 2225GMT are shown for measurements of spectral radiance temperature using the Barnes Model PRT-5 radiation thermometer. Interpretation of these observations without consideration of the solar irradiance, ground cover, and other canopy characteristics is not very meaningful. The purpose in presenting this figure is to illustrate the nature of the temperature profile observations that have been made.

The extent to which such information can be useful in thermal modelling studies is shown in Figure 2.3-11 and illustrates the variability of temperature between the three adjacent canopies during the day. The ordinate, maximum temperature difference, is defined as the magnitude of the difference between the highest and lowest temperatures measured by similar probes in the three different plots (Plots 52, 53, 54; 27 July 1975) tested as a function of time of day. The purpose in plotting the observations in this manner is to show what temperature variations are likely to occur between canopies with differing characteristics; that is, what kind of temperature differences are likely to exist at any one time under identical environmental conditions. The curve for probe number 1, soil surface temperature, indicates that at 2030GMT (1530 local time and approximately 90 minutes after solar noon) the maximum difference in soil temperature between the three plots amounts to 9.4^{o}C. For other probe locations the variability between the three plots is less in magnitude and shows less temporal variation. The exception is the temperature of the first green leaf which shows a very marked change in variability during the day.

These observations are still under analysis and some progress has been made in explaining the temperature profile in terms of the agronomic and environmental parameters. An extensive data archive has been organized that will permit ready access to the temperature profile measurements and other measurements made at the same time on these canopies.

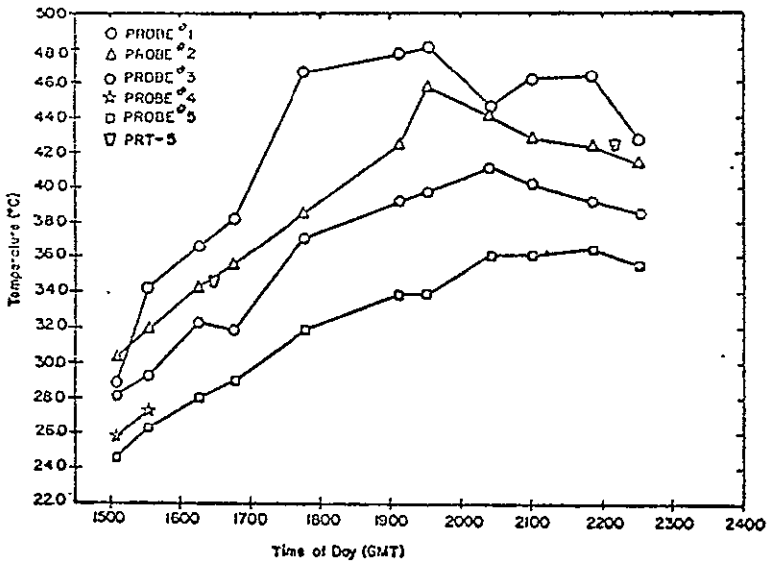


Figure 2.3-10. Temperature Profile in a Typical Mature Wheat Canopy (Plot 52, 27 July 1975)

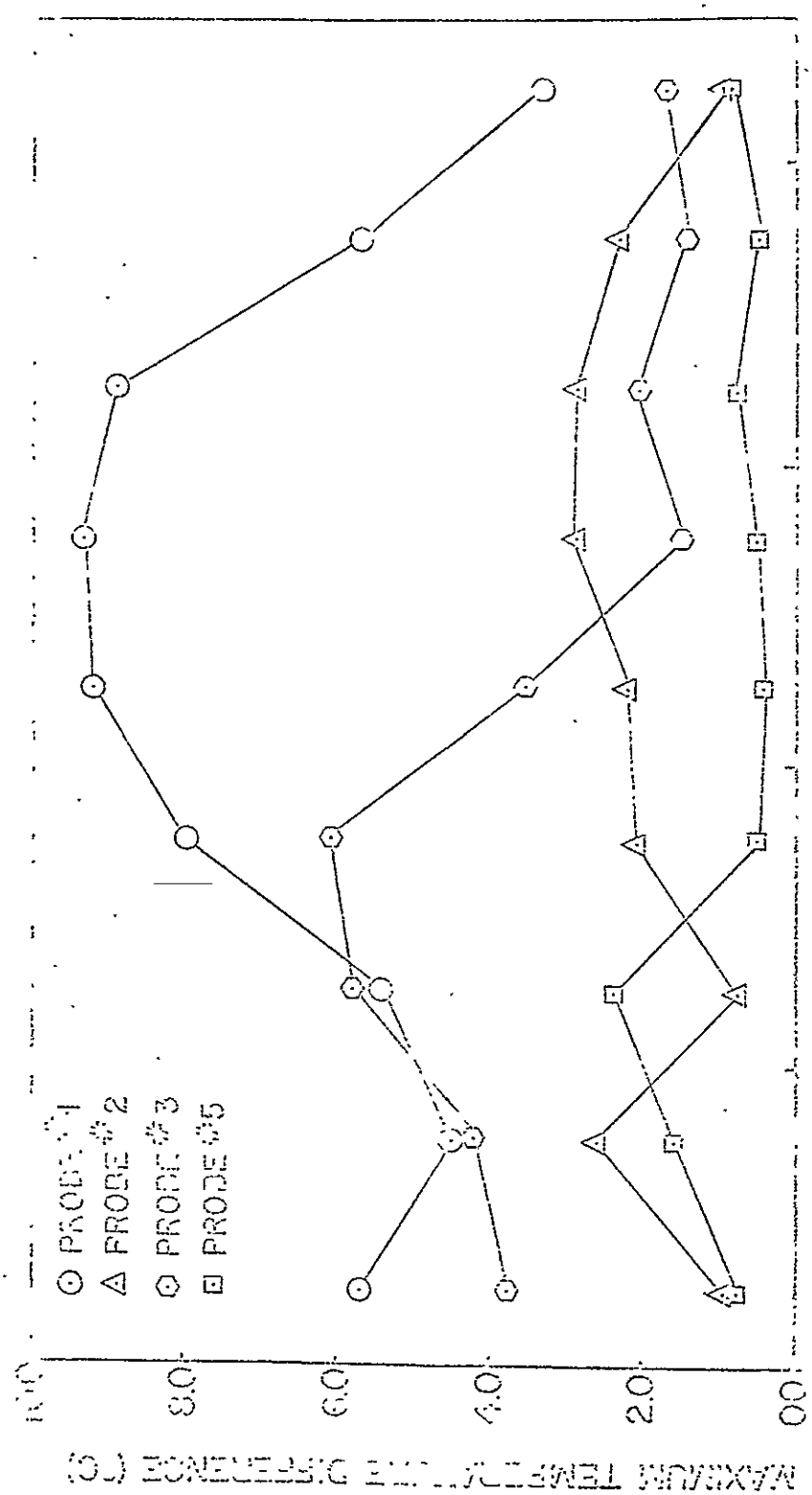


Figure 2.3-11. Variation of Canopy Temperature Profiles in Adjacent Test Plots (Plots 52, 53, and 54; 27 July 1976)

3. Spectral Radiance Temperature Measurements

The spectral radiance temperature, $T_R(\lambda)$, of a target is defined as the temperature of a blackbody having the same spectral radiance as the target. Using Wien's approximation to the Planck function, the relationship between the target temperature, T, and its spectral radiance temperature, $T_R(\lambda)$, is given as

$$\frac{1}{T} = \frac{1}{T_R} + \frac{\lambda}{C_2} \ln\varepsilon_\lambda$$

where the temperatures are in absolute units (K), $C_2 = 14{,}388\mu m \cdot K$ is the Second Radiation Constant, and ε_λ is the spectral emissivity. If the spectral bands over which the measurement is made is greater than say, 10% of the wavelength at the center of the band, then the derivation must be modified to properly define an effective wavelength at which the spectral radiance temperature is to be associated.

In the thermal infrared, the spectral emissivity of most natural targets is near unity but it should be quickly pointed out that a small variation in the emissivity has a large effect in the difference between the target temperature and the spectral radiance temperature as can be seen in Table 2.3-4. At the shorter wavelengths where the spectral emissivity may be in the 0.95 to 0.97 range for canopies and 0.85 to 0.90 for bare soils, the difference between radiometrically deduced temperatures, T_R, and temperature may differ by ½ to 5°C. In the longer wavelength region, like near 10μm which may approximate the effective wavelength of the Barnes Model PRT-5 radiation thermometer, variations in spectral emissivity from a value of unity are much less-suspected to be 0.97 to 0.99-for all types of targets; hence the difference between the temperatures so deduced would be quite small and probably less than 2°C.

An objective of the thermal modelling study is to relate the spectral radiance temperature-the parameter which is remotely sensed-to the temperature profile within the canopy as determined by contact sensors. In addition to the canopy emissivity effect discussed in the previous paragraph, it is necessary to consider what effect the temperature gradient within the canopy has on the remotely sensed spectral radiance temperature. This obviously is a complicated radiation transfer problem and the intent is to develop some understanding through parametric analysis much of which still remains to be performed.

The Figure 2.3- provides a comparison between the spectral radiance temperature (Barnes Model PRT-5, 8-13μm spectral region) and the canopy temperature profile. For the two times of day, 1650 and 2225GMT, the location in the canopy which most closely agrees with the spectral radiance temperature is that of the air just above the surface of the soil, Probe 2. Further analysis and additional temporal observations are necessary before this statement can be generalized. Particularly interesting will be to determine whether the most representative temperature is that of the soil surface air or the temperature of the air just above the canopy as is currently thought to be most important.

Table 2.3-4. The Difference Between Spectral Radiance Temperature $T_R(\lambda)$ and Temperature, T, as a Function of Spectral Emissivity and Wavelength for Target Temperature of 30°C.

Spectral Emissivity, ε_λ	$T_R(\lambda) - T$, K or °C		
	λ = 1μm	λ = 5μm	λ = 10μm
1.00	0.00	0.00	0.00
0.99	-0.06	-0.32	-0.64
0.98	-0.13	-0.64	-1.28
0.95	-0.33	-1.63	-3.26
0.90	-0.67	-3.36	-6.71
0.85	-1.04	-5.21	-10.4

A comparison of spectral radiance temperatures on the wheat canopies is possible by the analysis of spectral radiance observations from the Exotech spectroradiometer as shown in Figures 2.3-8 and 9. As discussed in the test associated with these figures, it was assumed the spectral radiance at 4.6μm and 11.25μm was due only to target emission and hence the spectral radiance temperature can be deduced as is reported in Table 2.3-5 for adjacent test Plots 52, 53, and 54 on 27 July 1976. Also on this same table are the Barnes PRT-5 observations (8-13μm band pass) and the spectral radiance temperature derived from appropriate averaging of the Exotech spectroradiometric observation over the 8 to 14μm spectral region. It should be noted these two results are within $\pm\frac{1}{2}^{o}$C agreement for this limited set of observations; this is to be expected by virtue of the instrument designs and calibration procedures. Further analysis on the variation of $T_R(\lambda)$ as a function of wavelength (including bandwidth), canopy characteristics, and temperature profiles are still in progress and it is not appropriate to make generalized conclusions from the limited observation sets examined.

Interception of Sunlight by Wheat Canopies

A new method utilizing a laser probe simulating sunlight has been developed and used to calculate the sites and magnitudes of energy interception by crop canopies. Initial results obtained for a spring wheat canopy are presented and discussed.

1. Description of Experiment

The geometrical characteristics of the wheat canopy in a commercial durum wheat field near the Agriculture Experiment Station were measured using an inexpensive laser on July 27, 1975 at Williston, North Dakota when the wheat was fully headed. Use of the technique allows the prediction of direct solar power and energy interception as a function of (1) leaves, stems, heads, awns, and soil, (2) depth into the canopy, and (3) time of day. The technique involves the use of the laser to statistically simmulate sunlight. The laser, mounted on a tripod, is aimed at the wheat canopy at a random zenith angle. The height above the ground at which the beam strikes an object in the canopy is recorded. Also recorded is the name of the object (i.e., leaf, stem, head, awn, or soil) and the zenith angle of the beam.

The accuracy of the estimate of the attenuation of the direct solar beam in the canopy is directly related to the number of measurements made. Two hundred points were measured in the modelling field. Data analysis involves averaging all points in a zenith angle 'window'. If there are 25 points in the window, then each laser hit is assigned a value of 0.04. One unit of sunlight is assumed to fall on the canopy at a zenith angle corresponding to the averaging window. As the beam front penetrates the canopy and passes the height of a laser hit, 0.04 is subtracted from the beam. Prediction of watts per unit area intercepted by each canopy component is accomplished by multiplying the normalized

Table 2.3-5. Spectral Radiance Temperature as a Function of Wavelength for Three Adjacent Test Plots (27 July 1975)

Plots	Time (GMT)	Spectral radiance temperatures, $T_R(\lambda)$, °C			
		λ = 4.6μm[1]	λ = 11.25μm[1]	λ = 8–14μm[1]	λ = 8–13μm[2]
52	1627	33.7	36.8	34.7	34.5
53	1625	33.2	34.0	33.1	33.9
54	1623	32.7	33.6	33.9	33.5
52	2209	41.5	42.2	41.9	42.0
53	2211	41.3	42.1	40.8	41.9
54	2214	42.0	42.7	42.5	42.6

Notes [1] From Exotech Model 100 Spectroradiometer observations

[2] From Barnes Model PRT-5 observations

intensity calculated from the heights of the laser hits by the magnitude of the solar flux. The magnitude of the solar flux must be measured experimentally.

2. Results

The interception of power by components in the canopy varied as a function of canopy height and of time during the day. Maximum power was intercepted near solar noon. This power was intercepted in the middle one-half of the canopy. Early in the morning and late in the afternoon all power was intercepted by the top most elements of the canopy. It is significant that very little power was intercepted by dead leaves near the ground. Photosynthetically active leaves intercepted direct solar power only during relatively high sun angles. The presence of the grain heads significantly diminished the power reaching photosynthetically active leaves which are located below the heads. Prior to heading the flag leaves of the canopy would intercept direct solar power throughout the day. The magnitude of the power intercepted by heads remained almost constant throughout the day.

Bidirectional Scattering Characteristics of Spring Wheat Canopies

1. Description of Experiment

The bidirectional scattering characteristics of a spring wheat canopy in the "modelling" field at Williston, North Dakota were measured on five dates during the 1975 growing season. On each date the LARS field spectroradiometer data acquisition system was used to obtain data continuously in wavelength, at eight azimuth and five zenith angles of view, and, weather permitting, at time intervals of one to two hours. Leaf area index, canopy biomass and other ground observation parameters describing the condition of the wheat canopy were measured on each date that spectral data were collected.

Analysis of data obtained on July 20, 1975 has involved display of the data graphically. The canopy bidirectional reflectance and two other variables were displayed on each graph. Implementation of the plotting technique required several hundred individual graphs, each graph displaying the data in a unique manner. The multidimensional character of the data required the large number of graphs.

2. Results

Analysis of the data using graphical techniques permitted the following conclusions to be drawn:

(1) The reflectance at nadir of the wheat canopy in the modelling field during the day of 20 July 1975 was not constant. The magnitude of reflectance at nadir increases with time during the day at all wavelengths. The increase during the day was about 100 percent in the red wavelengths and 30 percent in the near infrared.

(2) The reflectance of the wheat canopy is close to that of a lambertian reflector for data collected near solar noon in the visible region of the spectrum.

(3) The reflectance in the near infrared region of the spectrum is non-lambertian throughout the day.

(4) The deviation of the reflectance from a lambertian assumption is most pronounced at all wavelengths early in the morning and late in the evening. For an east-west azimuth angle the ratio of the reflectance at 60 degrees zenith angle to the reflectance at nadir is 2.6 at 23:11GMT.

(5) Except for two small regions of the spectrum, the spectral reflectance as a function of zenith and azimuth angles shows good correlation with the reflectance at nadir.

The measurements will be repeated during the 1976 growing season to verify the relationships identified in the first year's data. Certain improvements in the experimental procedure including sampling at more frequent time intervals, more replication of measurements, and more complete characterization of the physical attributes of the wheat canopy will be used during the second year.

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The activities of LARS in acquiring, processing, archiving, and analyzing field measurements data during the past contract year have been presented. Results describing the reflective and thermal properties of wheat canopies were presented. More extensive reports describing the analytical results obtained as well as several aspects of acquiring spectral measurements in field environments and quantitatively evaluating their quality will be published early next year.

A calibrated and fully annotated data containing spectral from several altitudes, agronomic, and meteorological measurements were obtained for winter wheat in Kansas and spring wheat in North Dakota during the past year. The data form one of the most complete remote sensing data sets ever acquired and should prove useful for research defining future remote sensing systems. Research results from the field measurements data are beginning to come forth and many studies using the data should be made in the next several years.

Experience gained during the first year of the project is now being applied to the acquisition and processing of data being acquired during the 1976 growing season. Procedures and systems developed and tested during the past year can be expected to result in smoother operation of the project, in improved data quality, and faster turn-around of data.

Specific recommendations for the field measurements effort during the third and subsequent years include:

- Acquisition of aircraft multispectral scanner data by a system with spectral bands from the entire optical portion of the spectrum.

- Implementation of data quality evaluation and verification procedures for all sensor systems.

- Additional efforts and resources devoted to characterization of crop and soil conditions at the time of each data collection mission.

- Development of plans for continuation of this type effort over new crops and locations. Specifically, the addition in 1977 of a test site for corn and soybeans in the Corn Belt is recommended. It is important to begin collecting data and gaining additional insights into the production of these crops prior to the expansion of LACIE to other crops.

ACKNOWLEDGEMENTS

The leader of the field measurements project at LARS is Marvin Bauer and the project manager is Larry Biehl. Barrett Robinson and William Simmons have been responsible for directing the acquisition and processing of data. Purdue faculty members, Leroy Silva, David DeWitt, and Byron Blair, have provided valuable guidance to the experimental work. Graduate students, Randolph Burch and Vern Vanderbilt have made major contributions to our knowledge of the reflective and thermal properties of crop canopies through their research. The efforts of David Freeman, Keith Phillip, William Zurney and Jim Kast in processing and reformatting data is gratefully acknowledged, as is the job Donna Scholz and Kathi Freeman have done in managing the data library. Other undergraduate and graduate students and staff working on the project have helped on many aspects of the project.

CONTENTS

2.4 Thematic Mapper Simulation

INTRODUCTION

An important question in remote sensing is "what is the optimum set of specifications for a multispectral scanner system?" The correct answer depends upon the index of performance selected as well as the class of applications for which the sensor system is to be optimized.

There are several ways to attack this question; one is empirically, i.e. using experimental data to simulate various sensor parameter combinations. It is the results of such a study which are to be reported in this section.

The ability to derive information from remotely sensed data gathered at a given time rests upon five classes of parametric values. These are:

1. The spatial resolution and scanning characteristics
2. The spectral sampling and bands used
3. The signal to noise characteristics
4. The amount of ancillary data available
5. The classes to be used, i.e. the particular information desired

It is especially important to note that these factors are interrelated to one another. Thus, assuming an empirical approach, the problem resolves itself to searching a five dimensional parameter space relative to the index of performance. It is obvious that this search cannot be done in an exhaustive fashion due to the size of a five parameter space, i.e., the number of possible combinations of the parameters. In this study the search was localized around the proposed thematic mapper parameters[1], a region suggested by the state of the art of constructing spaceborne multispectral scanners and the expected cost factors involved. Even so, it was necessary to limit the number of combinations tested. The scope of this investigation was primarily limited to three parameters - spatial resolution, noise level, and spectral bands although some variation in others was introduced.

There were two indices of performance used in this study. One was the accuracy achieved on multispectral pixels drawn from the central portions of agricultural fields. In this case emphasis is places upon the identification portion of the analysis task.

The second index of performance was the accuracy with which the correct areal proportions of each class in the flight line used could be estimated; this was done by determining the proportion of pixels assigned to each class by the classifier. In this case not only are "pure" pixels

from the central portion of agricultural fields involved, but so are mixed or multiclass pixels which overlap the field boundaries.

The general scheme of the study then was to simulate the desired parameter set by linearly combining the original pixels of the airborne data (IFOV≈6 meters) to form simulated pixels of the desired IFOV, then to classify this flight line using a machine implemented Gausian maximum likelihood pattern classification algorithm, and measure the index of performance.

Participants in the study included groups from NASA/JSC, NASA/ERL, GISS, NASA/Goddard and Purdue/LARS. After the participants as a group agreed upon the algorithms for producing the simulated data, simulation software was implemented at LARS. Simulation data sets were produced at the LARS Computation Facility and made available to analysts at JSC, ERL, GISS, Goddard and LARS. It was believed that variations resulting in analysis procedures and algorithms would enhance the value and credibility of the results. Indeed, this was the case as the results of the analysis relative to the simulated variables were similar and corroborated each other. Reported in this section is the work performed at LARS only.

Simulation Techniques

Data Utilized - Both airborne multispectral scanner data and field spectrometer data were used in this study. The multispectral scanner data was collected by the 24 channel MSDS system[2] aboard the NC-130 over the Finney County, Kansas, and Williams County, North Dakota, LACIE intensive test sites during the 1974-1975 growing season. Seven flight lines of the MSDS data were processed to simulate proposed thematic mapper spacecraft data (Table 2.4-1). The spectrometer data was collected by the NASA/JSC FSS S191H system, NASA/JSC FSAS interferometer system, and the Purdue/LARS Exotech 20C system over the Finney and Williams Counties intensive test sites and agricultural research farms. The spectrometer data also included the data collected over the Purdue Agronomy Farm by the Exotech 20C system during the summers of 1972, 1973, and 1974. A description of the three spectrometer systems can be found in the LACIE Field Measurements Project Plan. The spectrometer data were used to study wavelength band selections as well as to calibrate the airborne data for purposes of determining the signal to noise ratio.

Ground observations including crop types and field maps collected by USDA-ASCS personnel and color IR photography collected by the NC130 aircraft were used to support the multispectral scanner and spectrometer overflights. A discussion of the ground observations and descriptions of the two intensive test sites can be found in the LACIE Field Measurements Project Plan and in section 2.3 of this report. The ground observations were found to be fairly complete and very accurate.

Table 2.4-1. MSDS Data Selected for Simulation

Site	Date	Flight	Comments
* Williams County, N.D.	6/22/75	1	Poor crop calendar dat
Williams County, N.D.	6/22/75	3	Poor crop calendar dat
Finney County, KS	7/6/75	1	Serious banding
* Finney County, KS	7/6/75	3	Moderate banding
* Williams County, N.D.	8/15/75	1	Good set
Williams County, N.D.	8/15/75	2	Good set
Williams County, N.D.	8/15/75	3	Good set

* Those that were analyzed at Purdue/LARS

Preparation of Simulated Data - The simulation technique included spectral, spatial and radiometric considerations. A detailed discussion of spatial simulation algorithms is given in Appendix A. A general discussion of the simulation technique is given in this section.

a) Spectral Bands

MSDS system spectral bands which best matched proposed Thematic Mapper bands were selected. As seen in Table 2.4-2, the bands were relatively well matched. A combination of two MSDS bands was required to simulate one infrared band. The MSDS .53-.63 micrometer band was unavailable, and the .57-.63 micrometer band was substituted for simulation of the .52-.60 Thematic Mapper band. This substitution was recognized as suboptimal but the best alternative. The .74-.91 micrometer band was simulated as a result of speculation that Thematic Mapper bands 4 and 5 may be combined. A total of eight bands, therefore, were simulated.

Bands 4 and 5 were combined to form the .74-.91 micrometer band by equal weighted averaging after conversion to the reflectance domain. The combination of thermal bands was similarly achieved using the radiance domain.

b) Mean Angle Response Adjustment

A correction algorithm was applied to compensate for the non-uniform angular response characteristic due to the relatively wide view angle of the MSDS sensor. This effect is usually noted as one side of image nadir appearing brighter than the other side. The primary cause of the effect is that the scanner sees illuminated portions of the target at certain view angles and the shaded side at other angles. The correction made was to normalize the average scene response for each look angle. The normalization was made on each flight line independently. The method used for angle correction was:

- compute the average response of each look angle
- smooth the average to the least square error third order polynomial fit
- compute the inverse polynomial multiplicative correction function required to transform the polynomial curve to a constant value
- apply the correction function to each scan line

c) Spatial Degradtion

The spatial degradation procedure assumed a Gaussian total system modulation transfer funciton and compensated for aircraft scanner geometric distortions of unequal size and spacing of picture elements relative to scan look angle. Equations which define the method are given

Table 2.4-2. Correspondence of Thematic Mapper and MSDS Channels

Channel	Thematic Mapper	MSDS
1	.45 - .52μm	.46 - .50μm
2	.52 - .60	.57 - .63
3	.63 - .69	.64 - .68
4	.74 - .80	.76 - .80
5	.80 - .91	.82 - .87
6	1.55 -1.75	1.52 -1.73
7	10.4 - 12.5	10.0 - 11.0+11.0 - 12.0
8	.74 - .91	.76 - .80+.82 - .87

in Appendix A. Figure 2.5-1 shows a conceptual illustration of the simulated system spatial response. The simulated picture element in the scan line direction has a Gaussian point spread function with the IFOV specified as the distance between the half amplitude points; the Gaussian function is truncated beyond the 10 percent amplitude. In the along track direction, the point spread function is square. Using this definition of system point spread function, each degraded picture element was computed as the weighted average of higher resolution MSDS picture elements. Center-to-center spacing of degraded picture elements was equal to the width of the IFOV being simulated.

The aircraft scanner geometric distortion of unequal size and spacing of picture elements relative to scan look angle was accounted for. This "bow-tie" effect was factored into the computation of the weighting coefficients used to compute degraded pixel values. Since the coefficients used to counter the bow-tie effect were different for each simulated look angle, new weighting coefficients were calculated for each simulated pixel.

d) Reflectance Scaling

Calibration of MSDS data was required to allow combinations of reflective bands as discussed above and to appropriately scale the data according to specified LANDSAT dynamic range parameters, see Table 2.4-3[1]. To satisfy these requirements, calibration was achieved by:

- assuming the scanner black-body calibration source response corresponds to zero scene reflectance,
- assuming the scanner calibration lamp corresponds to a scene reflectance equal to the lamp equivalent reflectance,
- and using linear interpolation from these two calibration points.

Lamp equivalent reflectance data was supplied by JSC. It was well known that this method is not extremely accurate but it was believed to be adequate for purposes of band combination and dynamic range scaling.

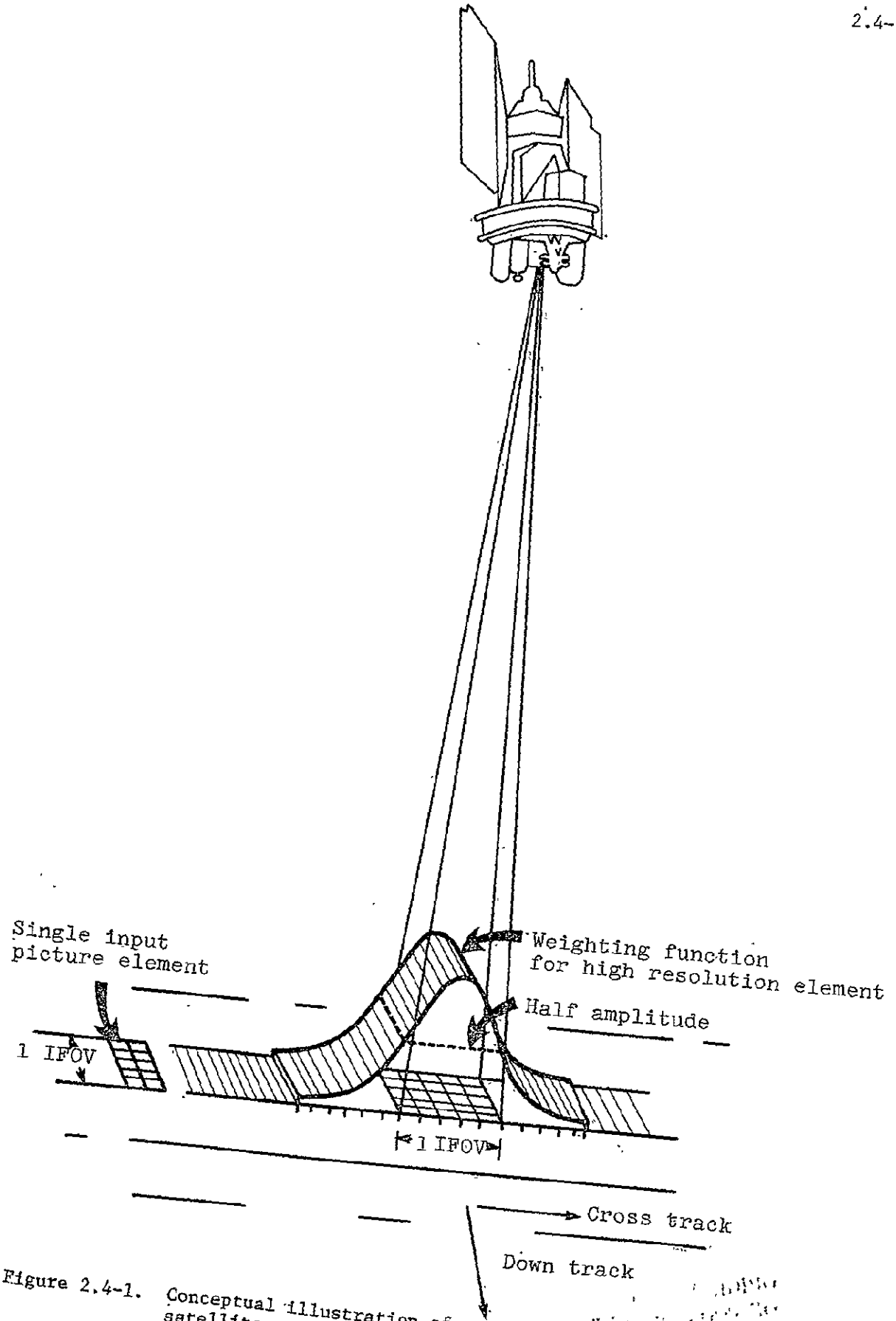


Figure 2.4-1. Conceptual illustration of a picture element viewed from the satellite.

Table 2.4-3. Thematic Mapper Parameters

Band (micrometers)	Saturation Surface Reflectance	Noise $NE\Delta\rho$ $NE\Delta T$	Spatial Resolution (meters)
.42 - .52	20%	.005	30 - 40
.52 - .60	58%	.005	30 - 40
.63 - .69	53%	.005	30 - 40
.74 - .80	75%	.005	30 - 40
.80 - .91	75%	.005	30 - 40
1.55 -1.75	50%	.005	30 - 40
10.4 -12.5	270 - 330K	.5K	90 -200
.74 - .91	75%	.005	30 - 40

e) Signal-to-Noise Degradation

Ideally, the degraded data would have a negligible noise level such that the impact of various noise levels on classification accuracy could be studied. In this case, however, the input data noise level was extremely high and the noise level after spatial degradation was approximately equal to the level specified for the LANDSAT-D system. To study the effect of additional noise, however, specific quantities of noise were added. Noise was added in the form of white Gaussian random numbers with standard deviation scaled to the desired noise equivalent reflectance. Additional details on noise considerations and calibration requirements necessitated are given in the section on evaluation of simulated data.

f) Implementation

The simulation process was implemented in four computer processing phases. The phases were:

- selection of spectral bands from the MSDS 24 channel computer data tape and reformatting the data into the LARSYS Version 3.1 data library.
- scan angle response normalization
- band combination, spatial degradation, and dynamic range adjustment
- addition of noise

Each phase produced separate outputs designed for analysis and processing in subsequent phases. Modularized processing had the advantage of providing data for analysis which had been processed in various stages of simulation. In addition, the method reduced processing redundancies. For example, a data set could be processed in phase four several times, each time adding a different level of noise, without the necessity of repeating previous processing steps.

A total of 62 data runs were prepared for analysis. Four spatial resolutions were simulated for three flights of four dates. Seven levels of noise were added to two flightlines. A list of data sets generated is shown in Table 2.4-4.

Evaluation of the Simulated Data - Examinations of the 24-channel multispectral scanner (MSDS) data revealed the presence of problems which resulted in limiting its usefulness. The true impact of the data problems were not known until simulation data were generated and classification results completed. Data quality problems present in the data are banding, bit errors, saturation, and inoperative bands.

Banding is evidenced in the imagery as alternating dark and light shading. The frequency of the banding varied from three to sixteen scan lines per band cycle for the flightlines considered. The banding can be seen in all MSDS reflective spectral channels (1-13) and channels 21 and 22 of the thermal data. Imagery illustrating the banding is shown in Figure 2.4-2. The banding signal amplitude was 1-15 data counts with larger amplitudes noted in spectral bands 1-8. It has been learned that the

Table 2.4-4. Simulation Data Sets Generated for Analysis

LARS Run	Test Site	Data Collected	Flight Line	Simulated Resolution*	Noise Aided NCAP
75002730	Williams	6/22/76	3	30	0
75002740	Williams	6/22/76	3	40	0
75002750	Williams	6/22/76	3	50	0
75002760	Williams	6/22/76	3	60	0
75001730	Williams	8/15/76	1	30	0
75001740	Williams	8/15/76	1	40	0
75001750	Williams	8/15/76	1	50	0
75001760	Williams	8/15/76	1	60	0
75003730	Finney	7/6/76	3	30	0
75003740	Finney	7/6/76	3	40	0
75003750	Finney	7/6/76	3	50	0
75003760	Finney	7/6/76	3	60	0
75002430	Williams	6/22/76	1	30	0
75002440	Williams	6/22/76	1	40	0
75002450	Williams	6/22/76	1	50	0
75002460	Williams	6/22/76	1	60	0
75001930	Williams	8/15/76	3	30	0
75001940	Williams	8/15/76	3	40	0
75001950	Williams	8/15/76	3	50	0
75001960	Williams	8/15/76	3	60	0
75003830	Finney	7/6/76	1	30	0
75003840	Finney	7/6/76	1	40	0
75003850	Finney	7/6/76	1	50	0
75003860	Finney	7/6/76	1	60	0
75001830	Williams	8/15/76	2	30	0
75001840	Williams	8/15/76	2	40	0
75001850	Williams	8/15/76	2	50	0
75001860	Williams	8/15/76	2	60	0
75001731 75001737	Williams	8/15/76	1	30	.0025, .005, .0075, .01, .015, .02, .03
75001741 75001747	Williams	8/15/76	1	40	.0025, .005, .0075, .01, .015, .02, .03
75003731 75003737	Finney	7/6/76	3	30	.0025, .005, .0075, .01, .015, .02, .03
75003741 75003747	Finney	7/6/76	3	40	.0025, .005, .0075, .01, .015, .02, .03

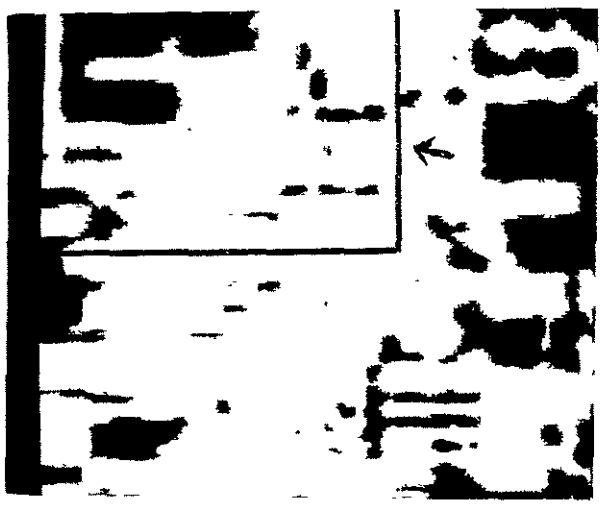
*Reflective data only, all thermal data at 120 meterm

.64-.68 micrometer band
6 meter IFOV

11.0-12.0 micrometer band
6 meter IFOV

Figure 2.4-2. Unprocessed MSDS data showing banding noise

.64-.68 micrometer band
30 meter IFOV



10.0-12.0 micrometer band
120 meter IFOV

Figure 2.4-3. Simulated data showing reduced banding noise

banding was probably caused by a loose mechanical joint within the detector housing and vibration of certain signal cables. The spatial degradation process significantly reduced the banding. Imagery illustrating the extent of banding noise reduction is shown in Figure 2.4-3.

System bit errors were noted in all spectral bands. This problem is illustrated in Figure 2.4-4. The histogram shows a much higher frequency of occurence of odd data counts than even. In addition, various higher order bit errors were indicated in various spectral bands. The bit errors tended to be masked by the spatial degradation process as illustrated in Figure 2.4-5; nevertheless, the impact on information content of the data must still be present.

Full scale saturation (data count 255) was noted in several spectral bands of several flightlines. Saturation occured not only for roof top and highway data, but also for agricultural areas which are of interest for analysis purposes. Saturated data points were omitted from all analysis since their values do not represent an accurate measure of relative scene radiance.

Sensor spectral bands 4 and 15 were inoperative for all data collection missions.

As discussed in another section, MSDS data were calibrated to reflectances using available saturation reflectance data with the assumptions of zero electronic offset and no atmospheric effects. This calibration was needed for band combination and range adjustments. After initial simulation data sets were generated and distributed for analysis, a refined calibration procedure was implemented. (The project time frame did not permit use of the refined procedure during the simulation processing.) The primary purpose in implementing the refined procedure was to enable an accurate determination of signal-to-noise levels and to enable accurate addition of noise for simulation of data with higher noise levels. The procedure followed in this refined procedure was as follows.

Near the time of a low altitude MSDS overflight, reflectivity spectra of five canvas calibration gray panels were determined by a truck mounted spectroradiometer system referenced to pressed barium sulfate powder. Gray panel reflectivities and MSDS response data were related through linear regression. The regression equation, transforming low altitude MSDS data to absolute scene reflectivity, was then used to compute mean reflectivities of agricultural fields within the low altitude flightline. Being clearly distinguishable in the high altitude MSDS data (actual panels were not), these large fields were then used as calibration panels were in the low altitude case. Field reflectivities were related to MSDS high altitude relative response data through linear regression yielding a linear function which could then be used to transform high altitude relative data to (absolute) reflectance.

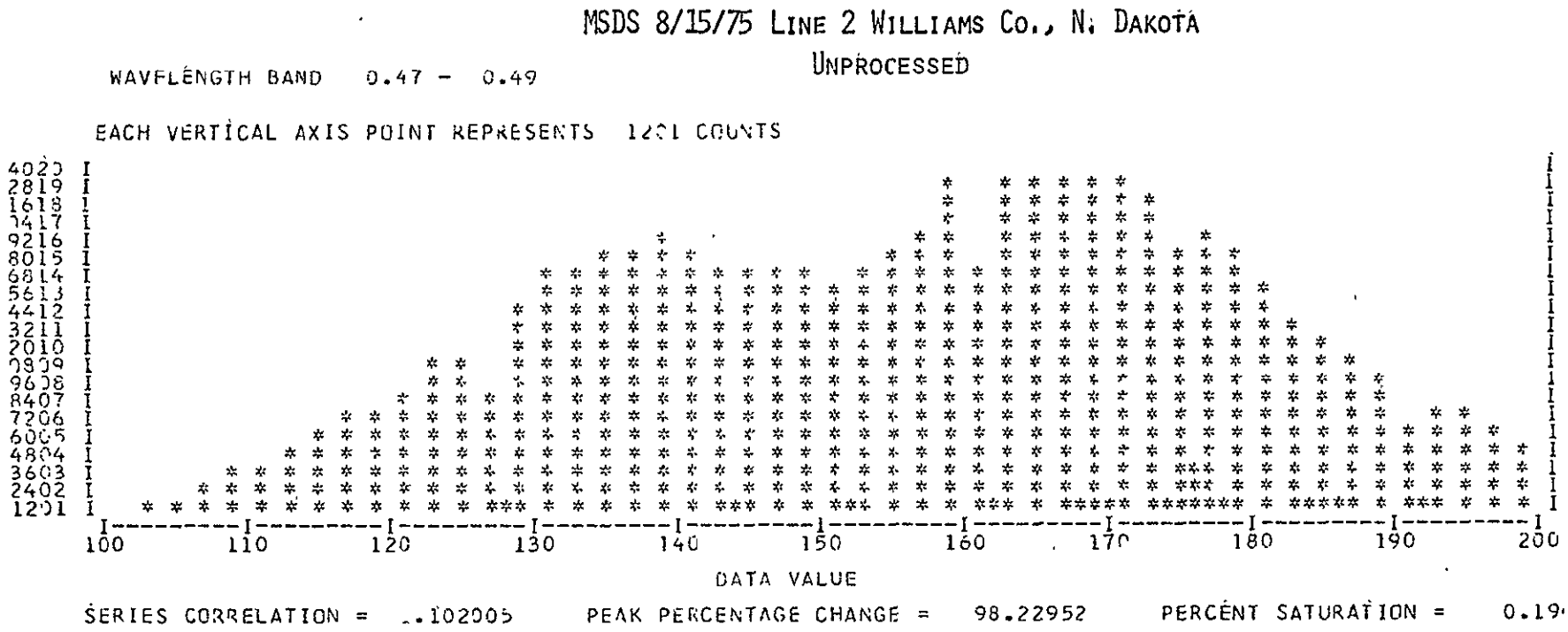


Figure 2.4-4 Histogram illustrating digitization bit errors.

MSDS 8/15/75 LINE 2 WILLIAMS CO., N. DAKOTA
SIMULATION PROCESSED

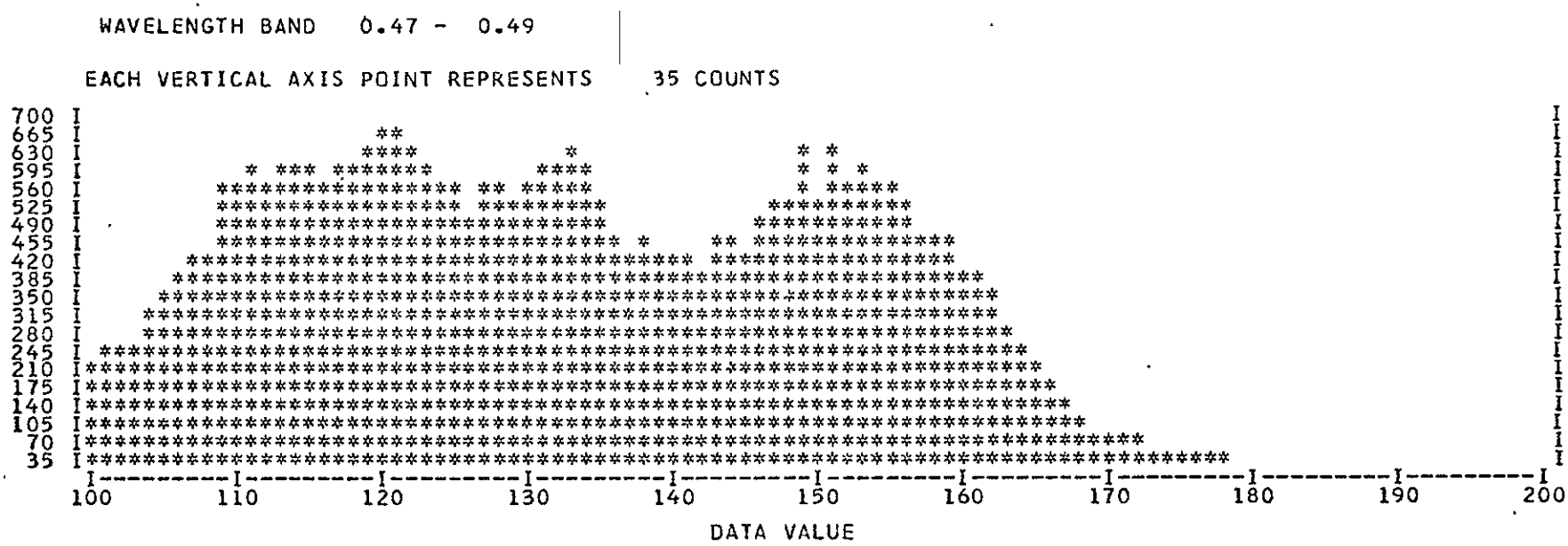


Figure 2.4-5 Histogram illustrating digitization bit error masked by the spatial degradation processing.

A least square error regression analysis was used which yielded the coefficients to the equation

$$y = A + Bx$$

For purposes of noise level computation, only the B term is needed. The noise level N is then

$$N = B\sigma$$

where B is the reflectance transform term from the regression analysis and the standard deviation of data values when scanning a constant target. Sigma was derived from the sixteen samples per scan line collected as the scanner low black body calibration source was viewed. The sixteen samples per scan line of calibration data were treated as an image and spectrally degraded in the same way as the ground scene data producing one calibration sample per simulated scan line. The standard deviation of the simulated calibration samples multiplied by B is the measure of noise used. Table 2.4-5 shows noise levels derived in this fashion. In addition, the B term was used to determine the magnitude of random numbers required in order to add a specified level of noise to the data. The standard deviation, σ, of the random numbers required to add a $NE\Delta\rho$ of Na is

$$\sigma n = \frac{Na}{B}$$

And finally, to test the functioning of the scan angle response normalization algorithm, the 30 meter data from flight line 1, ND, Aug. 15, 1975, was analyzed with a procedure intended to find if any effects of sun or scanner angle could be seen in the classification. The flight line was divided lengthwise into thirds and training fields were taken from each third. The three training sets were compared in the SEPARABILITY processor and no apparent differences due to location across the flightline could be seen. The flight line was also classified with the combined training sets and again no differences associated with training set locations were found in the classification.

Analysis Procedures Used

Each analyst was allowed some freedom in the training set selection but the procedures used did not differ greatly. Each analyst selected areas from which training fields were taken. Two analysts elected to divide the flightline into one mile wide strips across the full width and train on alternating strips. The other individual trained on alternating one mile sections. In each case, then, training sets were taken from one half of the area and were distributed systematically over the entire flightline. These candidate training areas were clustered primarily for image enhancement of the field boundaries so that the training fields could be more easily selected. There was, however, an

Table 2.4-5. Noise Levels of Simulation Data Before Adding Noise

Wavelength Band	Williams County August 15, 1975 Flight 1 NEΔρ 30 Meter Resolution	Williams County August 15, 1975 Flight 1 NEΔρ 40 Meter Resolution	Finney County July 6, 1975 Flight 3 NEΔρ 30 Meter Resolution	Finney County July 6, 1975 Flight 3 NEΔρ 40 Meter Resolution
.46 - .50	.001	.001	.003	.002
.57 - .63	.002	.002	.009	.005
.64 - .68	.006	.005	.017	.010
.76 - .80	.008	.008	.022	.012
.82 - .87	.006	.005	.004	.003
1.52 - 1.73	.003	.003	.023	.013
.76 - .87	.-04	.004	.009	.005

additional effect from clustering. This was the definition of spectral subclasses within fields and when subclasses were found the analyst could adjust the training sets to sample them.

Color infrared photographic mosaic prints were made from photographic data collected concurrently with the scanner data. Informational class information provided by ground observations was transfered to clear plastic overlays on the mosiac print. The analyst could then easily locate the corresponding fields in his cluster maps and assign the field coordinates to the informational classes.

Statistics were calculated for each training area and compared using the SEPARABILITY processor. Similar classes were combined, where indicated, and the data set was used to classify the flightline. Training areas were included in the test fields but actual field boundaries did not necessarily correspond between the training and test fields since the test fields had been pre-selected for the entire flightline.

The entire training set selection procedure was repeated for each resolution size so that any effects on training set selection which might be caused by data point resolution size would be included in the analysis results. An example is the increasing difficulty and eventual impossibility of selecting samples from small, or narrow, fields as the resolution size increases.

The two indices of performance previously mentioned were each applied in two ways. Classification (identification) accuracy was evaluated using both training and test sample performance while proportion estimation (identification and mensuration) was carried out over the flightline as a whole and as an average of portions of the flightline. Further details on each of these is as follows.

The training performance is the overall classification accuracy (number of training pixels correctly classified divided by the total number of training pixels) of the pixels used to calculate the individual class statistics. The test performance is the overall classification accuracy of the test field pixels. The test fields were selected in the original six meter data by choosing the largest rectangular block of pixels that would fit within the agricultural field so that no boundary pixels were included. The test field boundaries were then found in the degraded spatial resolution such that no "super" pixels (degraded spatial resolutions) containing boundaries were included. Some of the original test fields were discarded in this process because they became too small, i.e. there were no pure field center "super" pixels.

The RMS error of informational class proportion estimates for the flightline was found by calculating the percent of the flightline classified as a particular class and comparing it with the ASCS ground collected estimate using equation (1).

$$\text{RMS Error} = \sqrt{\frac{\sum_{i-1}^{N} (C_i - C'_i)^2}{N}} \qquad (1)$$

where N = number of informational classes

C_i = percent classified as informational class i

C'_i = percent of class i estimated from ASCS ground collected data

The fourth criterion stated above was found using equation (1) above on a section basis (approximately one square mile) and finding the average RMS error of all the sections in the flightline. Each flightline analyzed included a two by six mile area; so there were twelve sections in each flightline. The informational classes used for the three flightlines are given in Table 2.4-6.

LANDSAT 2 data for two of the flightlines were also analyzed, but test fields were not selected for the LANDSAT data.

Results

Spatial Resolution Parameter - The results from classification for the various IFOV's simulated indicate no significant trend in the training performance across the four resolutions (Figure 2.4-6 and Table 2.4-6). The training performances were all high, above 93%. The most change occurred in the 6/22 flightline which was probably due to the difficulty caused by the poor crop calendar date mentioned above.

The test performance results, however, indicate a general upward trend as the IFOV increases from 30 meters to 60 meters for two of the flightlines (see Figure 2.4-7 and Table 2.4-6). The third flightline shows a downward trend which again is probably caused by the difficulty due to the stage of growth of the crops. The upward trend in the test performance of the two flightlines is presumed to be caused by the better signal to noise ratio in the larger instantaneous field of view (IFOV) data. A 60 meter pixel is simulated by averaging approximately 100 six meter pixels as compared to approximately 25 for a 30 meter pixel.

The number of test fields varied in inverse relation to the IFOV, since as the IFOV increased, the probability that some test fields would not contain any pure pixels increased. To determine if this situation rather than improved signal to noise ratio might have caused the upward trend in the classification performance for larger IFOV's a common set of test fields were selected to test the performance of the classifications. The test performance increased slightly, 0 to .7%, however, the trend was the same.

Table 2.4-6. Informational Classes Used in the Analysis

Williams County, N. Dakota		Finney County, Kansas
6/22/75	8/15/75	7/6/75
Bare Soil	Harvested Wheat	Harvested Wheat
Grasses/Pasture	Unharvested Wheat	Corn
Small Grain	Grasses/Pasture	Grain Sorghum
	Fallow	Grasses/Pasture
	Other (corn-oats)	Fallow

Spatial Resolution Parameter
OverallTraining Performance
channels: 2,3,4,5,6,7
noise: .002-.008 NEΔP, .09 NEΔT for 30m 8/15-1
.004-.023 NEΔP, NEΔT for 30m 7/6-3

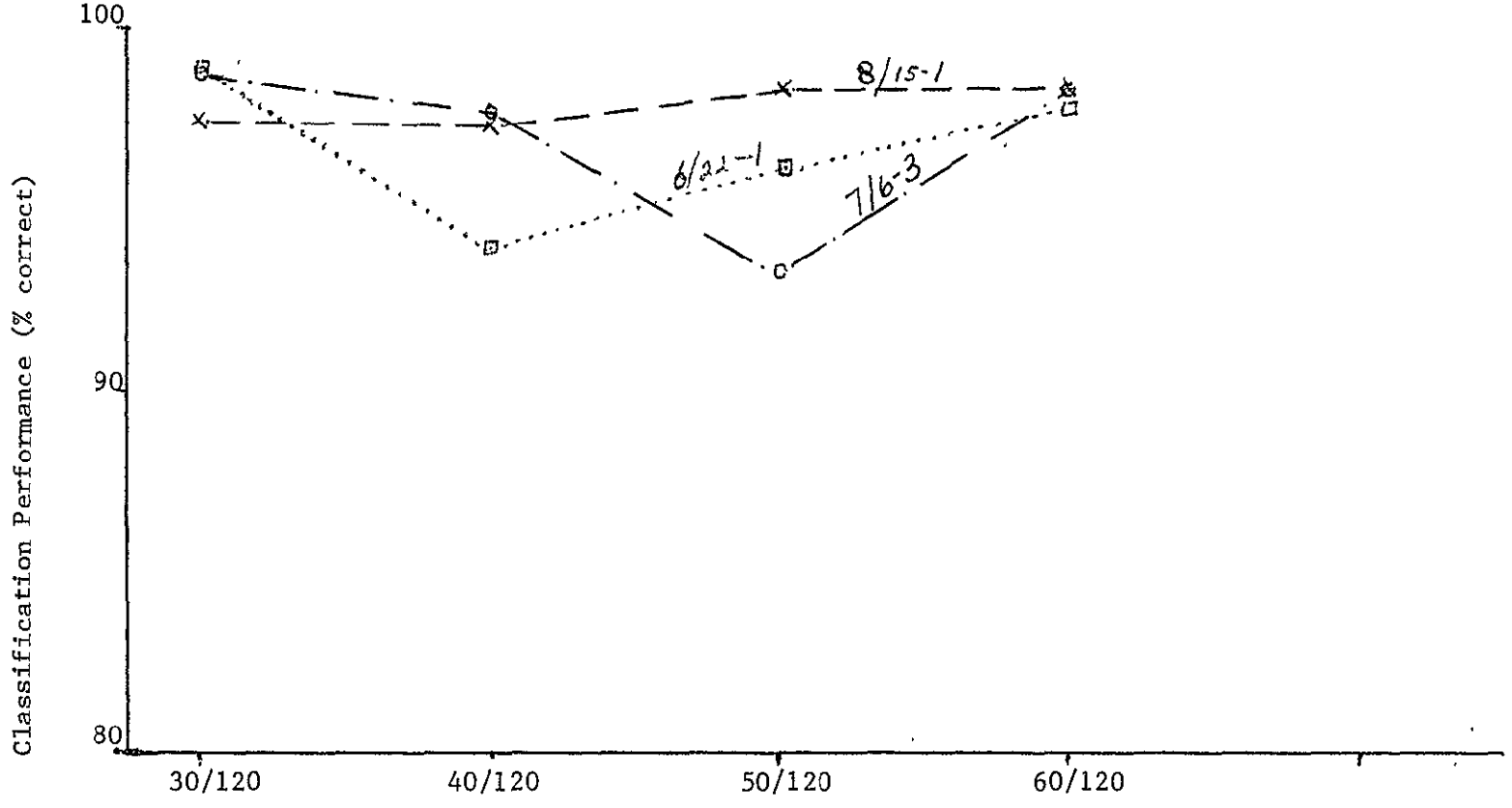


Figure 2.4-6 Reflective/Emissive IFOV (m)

Spatial Resolution Parameter
Overall Test Performance
Channels: 2,3,4,5,6,7.
noise: .002-.008 NEΔP, .09 NEΔT for 30m 8/15-1
.004-.023 NEΔP, for 30m 7/6-3

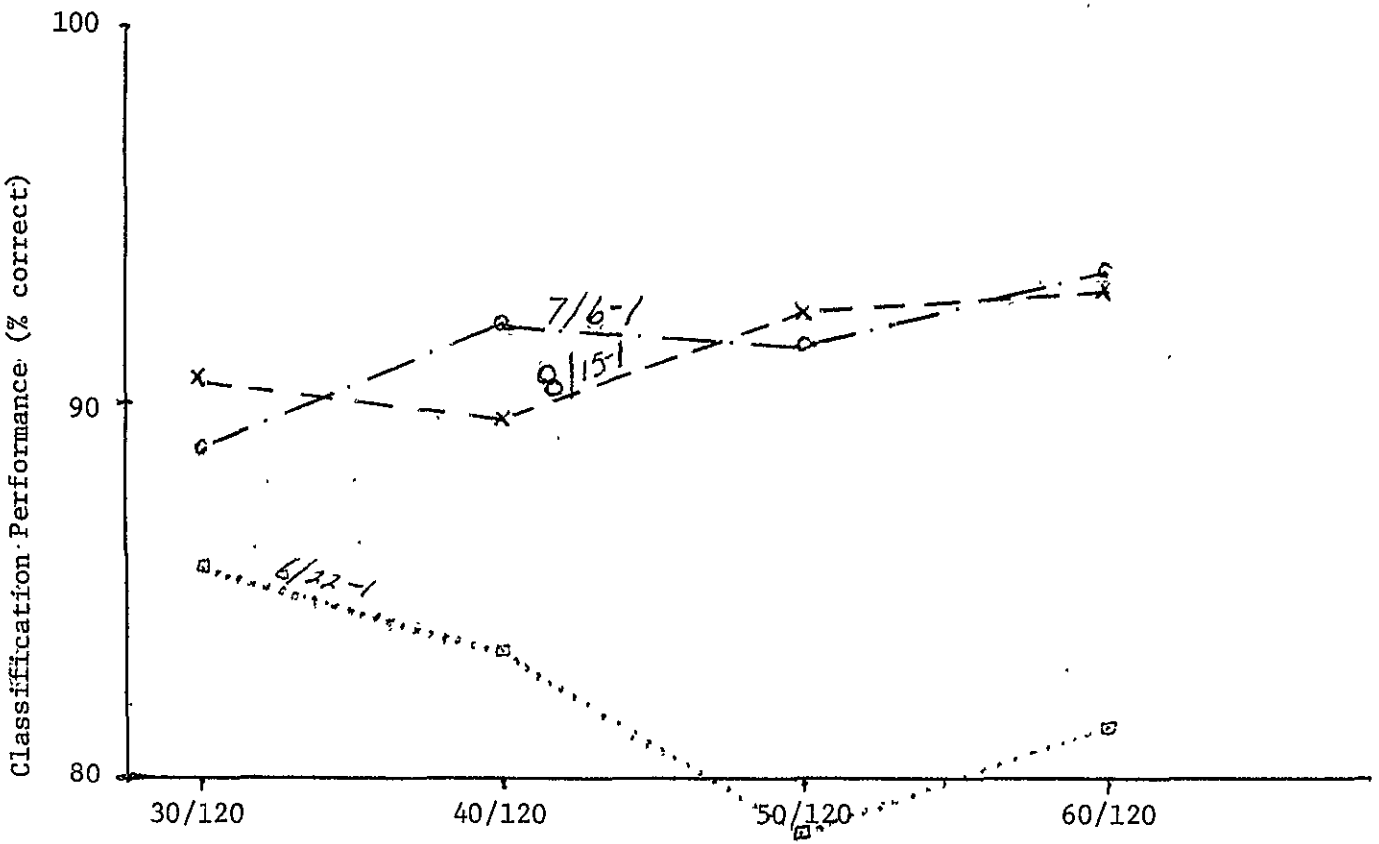


Figure 2.4-7 Reflective/Emissive IFOV (m)

The decreased percentage of field center pixels as the IFOV increases can be noted from the percent of the flightlines used for test (see Table 2.4-7). For each of the three data sets the percentage of field center pixels used for testing dropped an average of 10 percent from 30 to 60 meters.

Table 2.4-7 also indicates that a significantly higher proportion of the Kansas flightline was used for testing than the North Dakota flightline. This is due to the larger field size in Kansas (see Table 2.4-8). The differences between the 6/22 date and 8/15 date over the North Dakota flightline are due to the exclusions of anomolies (bare spots) in the centers of fields which were present in the 8/15 data and not in 6/22 data. The rectangular test fields were reduced so that the bare spots would not be included in the test performance.

The RMS error of the proportion estimates for the flightline and the average of twelve sections (see Table 2.4-6 and Figures 2.4-8 and 2.4-9) indicate that the least error is obtained using a 30 meter spatial resolution. The RMS errors increased as the IFOV increased from 30 meters to 40 meters to 50 meters and then leveled off or dropped slightly as the spatial resolution increased from 50 meters to 60 meters. The two RMS errors increased again for the Landsat 2 data. The Landsat 1 and 2 data have an IFOV of approximately 90 meters if the definition of spatial resolution being applied for the Thematic Mapper is used; the spatial resolution of Landsat 1 and 2 more commonly known as 80 meters.

The reduced error in the proportion estimates as the IFOV is reduced is due to the increased percentage of pure field center pixels to boundary pixels. The test performance criterion was based only on the pure field center pixels. Many errors occur in delineating boundary pixels because very often they aren't similar to either of the classes that they represent. For example, a pixel including both bare soil and wheat may appear similar to grass.

The two criteria using RMS errors include the boundary errors, however, it should be noted that they aren't a direct measure of boundary errors. It is possible for the boundary errors to cancel each other out over a given area so that the proportions estimates obtained from the classifications are more nearly correct. This is the reason that the RMS error in the proportion estimates for the entire flight are smaller than the average RMS errors of the twelve sections in the flightlines for a given resolution. The RMS errors, however, are the better measures in comparing the differences of spatial resolutions because these criteria are based on all pixels in the flightline. Better area estimates are possible with the smaller of the five IFOV's.

An analysis of variance was run on the RMS errors for the four resolutions using the RMS errors for the twelve sections in each of the three flightlines to determine if the differences were significant. A partially nested design with equal cell sizes from the BMD Biomedical

Table 2.4-7. Percentage of Test Sites Used for Training and Test

North Dakota - 6/22

Performance Criteria	Percentage of Test Site for Given Resolution (m)			
	30	40	50	60
Train	12	15	17	16
Test	38	35	31	28

Kansas 7/6

Train	25	35	36	37
Test	55	50	46	42

North Dakota - 8/15

Train	19	21	20	21
Test	27	24	23	19

Table 2.4-8. Characteristics of the Agricultural Fields by Flightlines

Location	Date	Flightline	No. Fields	Ave. Field Size (acres)	Field Size Range (acres)
N. Dakota	6/22/75	#1	250	29.3	1-480
Kansas	7/6/75	#3	187	39.3	1-161
N. Dakota	8/15/75	#1	250	29.3	1-480

Table 2.4-9. Analysis of Variance Results for Spatial Resolution

Source	Sum of Squares	Deg. of Freedom	Mean Square
I - resolution	2.16	3	.72
J - flightline	8.89	2	4.45
K(J) - sections	32.38	33	.98
IJ	1.02	6	.17
IK(J)	15.24	99	.15

F value {I/IK(J)} = 4.67 $\quad F_{.95}(3,99) \approx 276$

Spatial Resolution Parameter
RMS Error of Preportion Estimate for Flight Line
channels: 2,3,4,5,6,7
noise: .002-.008 NEΔP, .09 NEΔT for 30 m 8/15-1
.004-.023 NEΔP, for 30 m 7/6-3

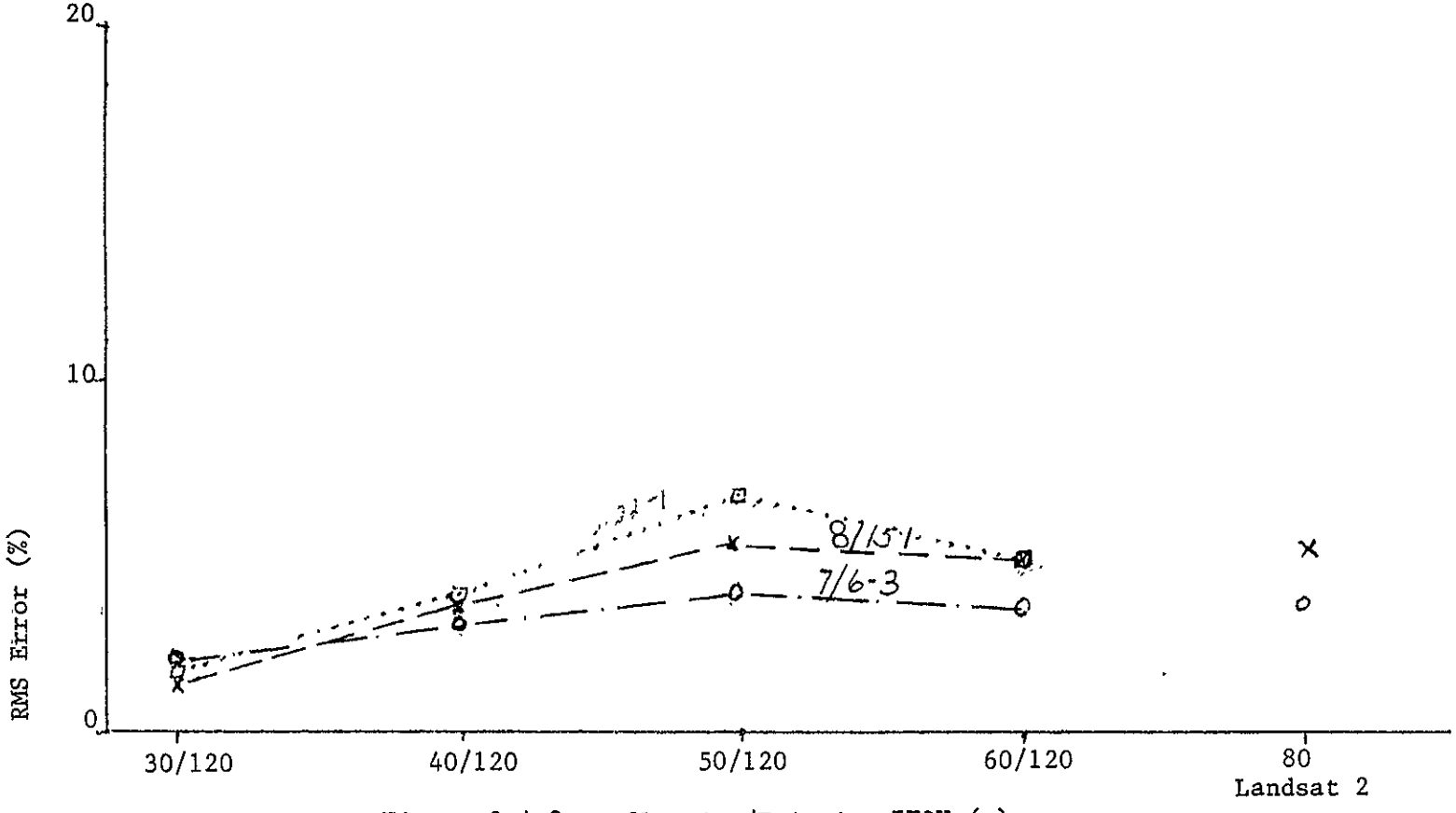


Figure 2.4-8 Reflective/Emissive IFOV (m)

Spatial Resolution Parameter
Average RMS Error of Proportion Estimates of Sections in Flight Line
channels: 2,3,4,5,6,7
noise: .002 - .008 NEΔP, .09 NEΔP for 30m 8/15-1
.004 - .023 NEΔP, for 30m 7/6-3

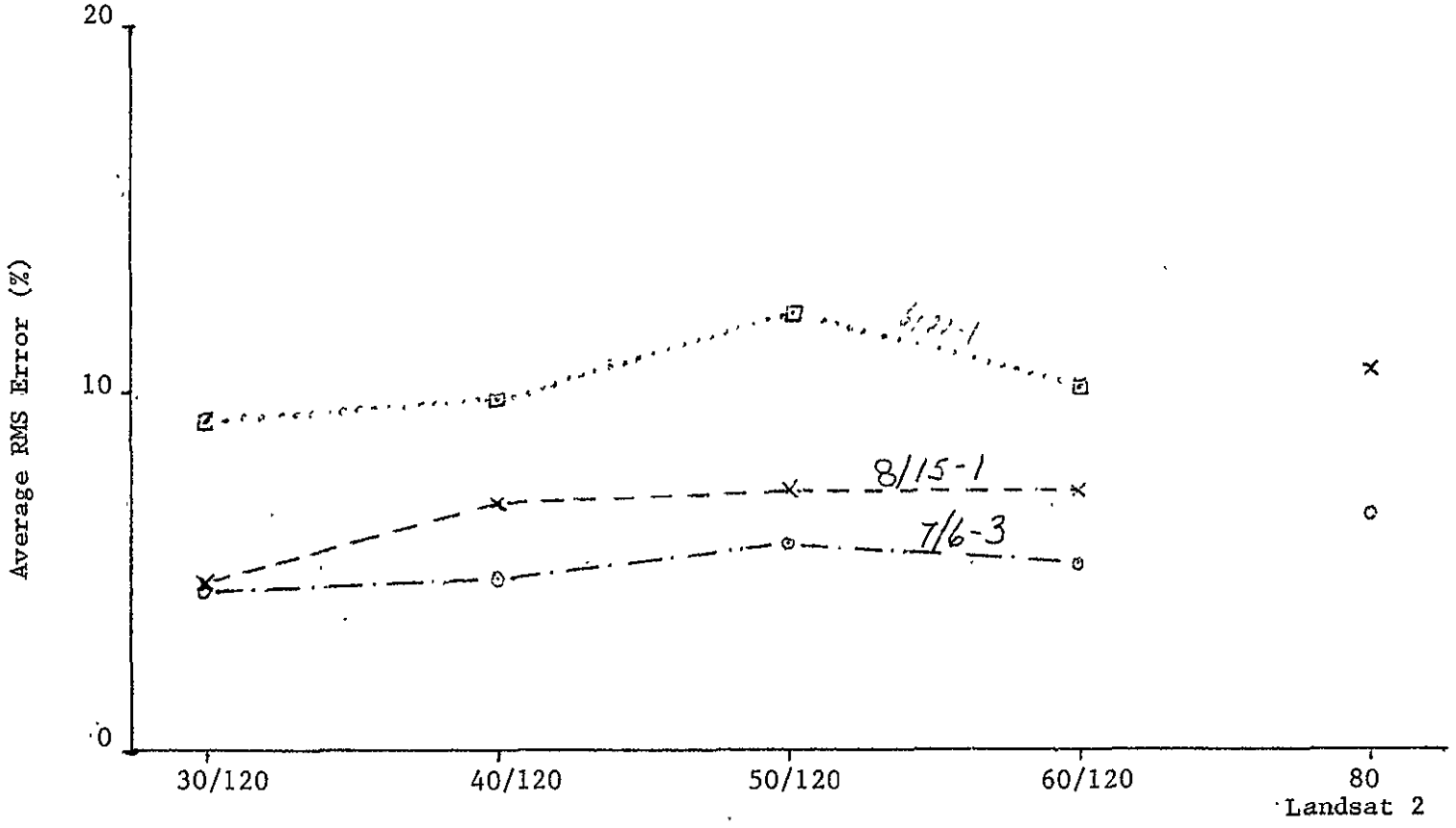


Figure 2.4-9 Reflective / Emissive IFOV (m)

Computer Programs was used (BMD08U)[3]. The differences in the RMS errors for the four resolutions were significant at the .05 significance level (see Table 2.4-9).

Noise Level Parameter - The analysis technique for the noise level parameter included using the training fields selected in the no noise added case and re-estimating the multivariate Gaussian statistics in each of the seven noise-added data sets for a particular IFOV and flightline. Each of the noise added data sets were then classified using simulated Thematic Mapper channels 2, 3, 4, 5, 6, and 7. The evaluation criteria of the noise level parameter include the overall training performance and the overall test performance. See Figures 2.4-10 thru 2.4-12 and Table 2.4-10.

To actually simulate different noise levels in data from satellites, the variable of analyst's determination of field boundaries might have been included. This would have necessitated the time consuming routine of the selection of training areas from each noise added data set independently from the other sets. Tests were run which illustrated that the data clustered nearly the same for the noise levels of .0025 to .015 NEΔP added. The field boundaries, however, were difficult to distinguish in the cluster maps for the noise levels of .02 and .03 NEΔP added levels. In light of these results the performances found for the .02 and .03 NEΔP noise added levels may be optimistic, since the field boundary delineation difficulty is not included.

The original plans were to simulate the .005 NEΔP noise level planned for the Thematic Mapper for all channels (also .01 NEΔP for channel 6) together with .5, 1.3, 1.6, 2 and 3 times that noise level. The noise level present in the original MSDS data, however, was too high for the original plans. After the averaging to simulate 30 to 60 meters, the noise level for the channels were of the same magnitude as planned for the Thematic Mapper - the 0 added noise case. To simulate higher levels .0025, .005, .0075, .01, .015, .02, and .03 NEΔP (x100 for NEΔT) noise levels were added to the 0 added set. The calibration for the noise addition was obtained using the grey panels at the calibration location in the intensive test site and the truck mounted spectrometric data is previously described.

In each of the four data sets analyzed across all eight noise levels, once the level of noise added became greater than the noise already in the data, the train and test performances fell off significantly. It is difficult, however, to draw any conclusions relative to the Thematic Mapper since the noise levels were not constant across all bands.

Spectral Band Parameter - The first analysis technique for the wavelength band set as the parameter consisted of selecting training areas to represent spectral classes from cluster maps obtained using simulated Thematic Mapper channels - 2, 3, 4, 5, 6 and 7 - the same technique as used for the spatial resolution parameter. The simulated

Noise Parameter
Overall Train Performance
Channels: 2,3,4,5,6,7
Resolution: 30/120 meters
Noise: for 0 added .002-.008 NEΔP, .09 NEΔT 8/15-1
.004-.023 NEΔP, NEΔT 9/6-3

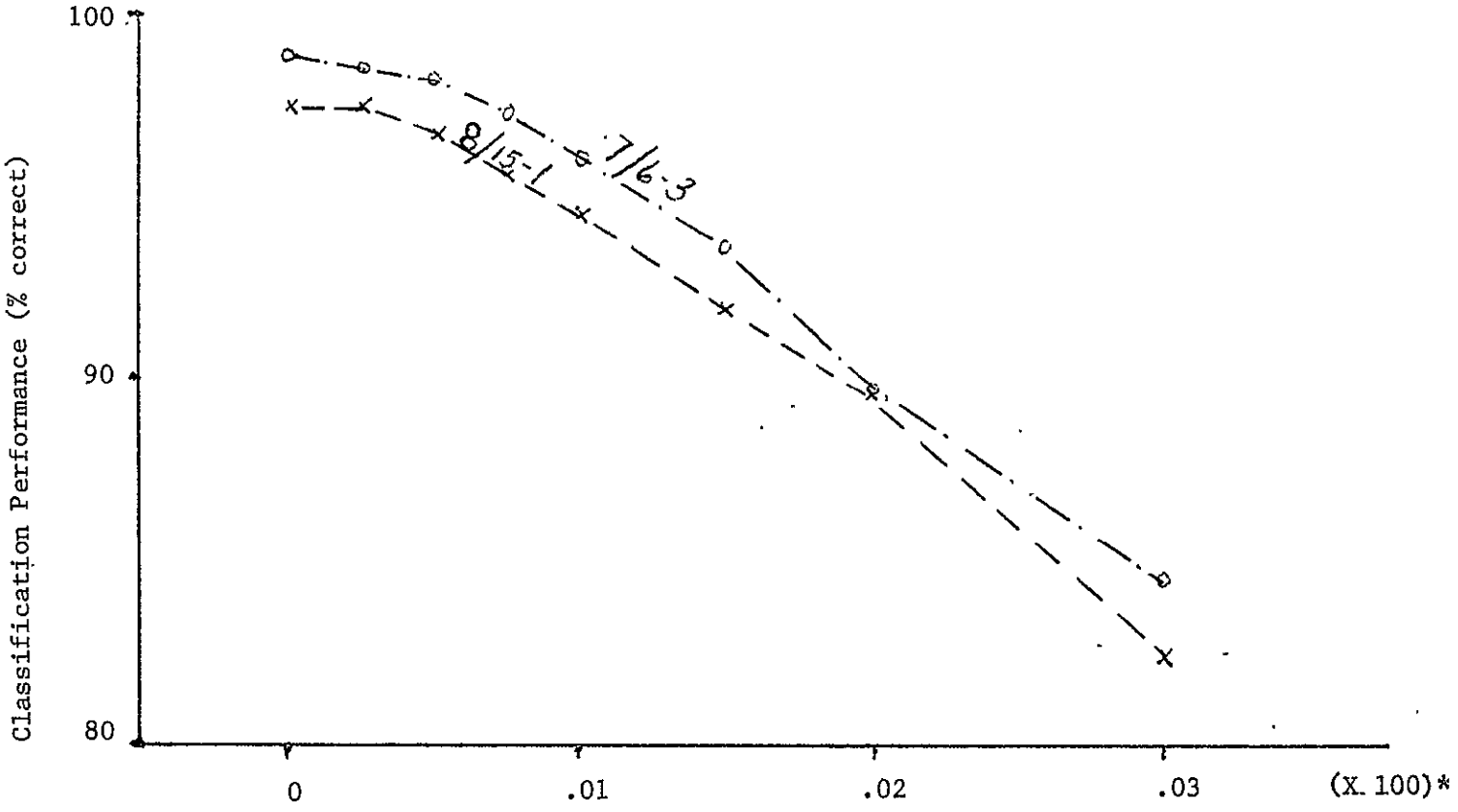


Figure 2.4-10 Noise added (NEΔP or* NEΔT)

Noise Parameter
Overall Test Performance
Channels: 2,3,4,5,6,7
Resolution: 30/120 meters
Noise: for 0 added .002-.008 NEΔP, .09 NEΔT 8/15-1
.004-.023 NEΔP, NEΔT 7/6-3

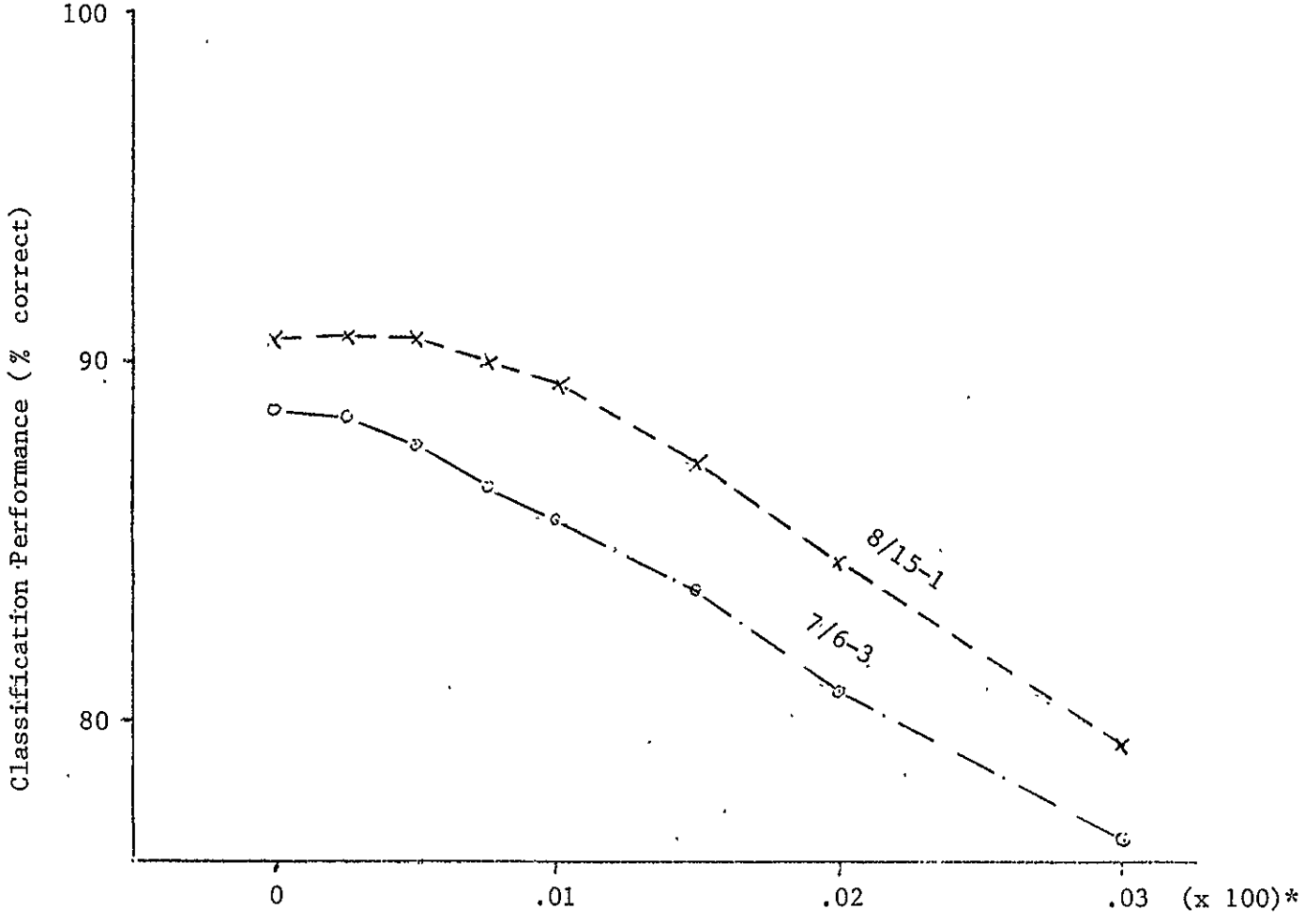


Figure 2.4-11 Noise Added (NEΔP or *NEΔT)

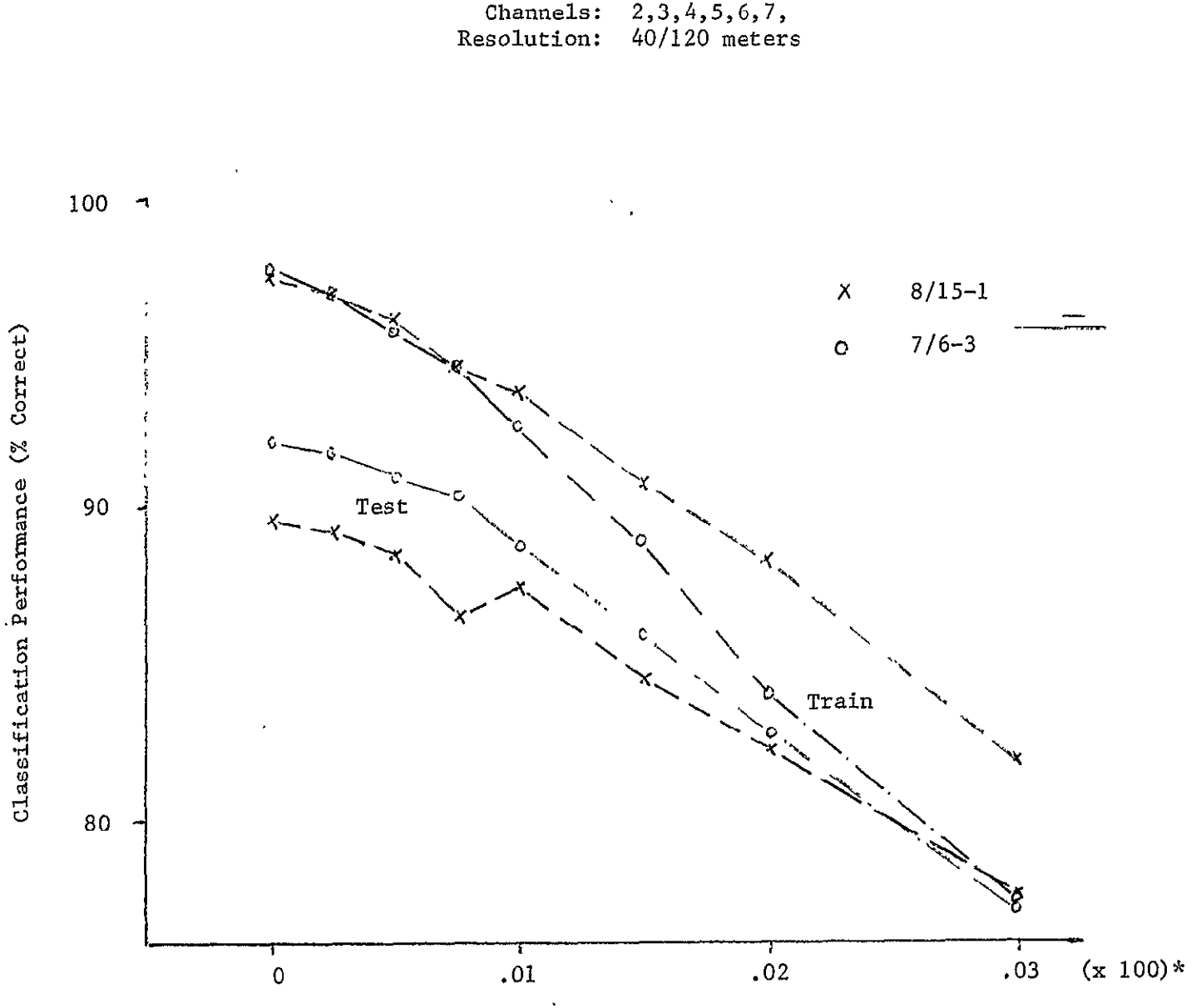


Figure 2.4-12 Noise Added (NEΔP or *NEΔT)

Table 2.4-10

Results of Noise Level Parameter Study

Spatial Resolution - 30/120 meters

Overall train performance (per cent correct)

Noise level added to original data (NEΔP or *NEΔT)

Site - Date	0	.0025	.0050	.0075	.010	.015	.020	.030
Kansas - 7/6	98.8	98.5	98.2	97.2	96.0	93.6	89.7	84.6
N. Dakota - 8/15	97.5	97.4	96.7	95.6	94.5	91.9	89.6	82.5
Overall test performance (per cent correct)								
Kansas - 7/6	88.8	88.5	87.8	86.6	85.7	83.7	80.9	76.7
N. Dakota - 8/15	90.7	90.8	90.7	90.1	89.5	87.3	84.5	79.4

Spatial Resolution - 40/120 meters

Overall train performance (per cent correct)

Kansas - 7/6	97.7	97.0	95.6	94.5	92.5	88.9	84.0	77.6
N. Dakota - 8/15	97.3	96.9	96.0	94.5	93.7	90.8	88.3	81.9
Overall test performance (per cent correct)								
Kansas - 7/6	92.1	91.8	91.0	90.4	88.8	85.9	82.7	77.2
N. Dakota - 8/15	89.6	89.3	88.5	86.5	87.5	84.1	82.3	77.7

NEΔT is 100 times value given.

30 meter North Dakota - 8/15 and Kansas - 7/6 flightlines were then classified using four different feature sets. The results are given in Table 2.4-11 and Figures 2.4-13 and 2.4-14.

The four feature sets selected resulted from discussions at NASA/JSC concerning possible ways to reduce the number of proposed channels to six and attempts to grossly simulate the present Landsat 1 and 2 scanners. Channels 1, 2, 3, 4, 5, 6 and 7 represent the originally proposed Thematic Mapper channels[1]. Channels subsets 2, 3, 4, 5, 6 and 7 and 1, 2, 3, 6, 7 and 8 represent frequently discussed combinations of six channels. Channel 8 is the combination of channels 4 and 5. Feature set 2, 3, 4 and 5 grossly approximately the same spectral range as the present Landsat 1 and 2 scanners cover.

The results indicate that slightly higher performances are possible for these data sets when Thematic Mapper channel 1 is included. The results also indicate little or no change in performance if Thematic Mapper channels 4 and 5 are combined into one channel. The differences are very small, however. An analysis of variance has not yet been run to determine if the differences were significant. It is possible that the previously described unusual noise in the MSDS data is acting to minimize any significant differences due to spectral band changes.

A second analysis technique relative to spectral band selection is to study the correlation of the proposed Thematic Mapper (T.M.) channels. This was done for the agricultural crops in the intensive test sites and at the Purdue Agronomy Farm, sampling the growing season using spectrometer data. The correlation of the proposed Thematic Mapper band .52-.60μm and the MSDS substituted band .57-.63μm was also studied. Concern had been raised that the .57-.63μm channel of the MSDS data did not represent the .52-.60μm Thematic Mapper channel well. Also included in the correlation study were the Skylab S192 scanner bands which cover the range between 1.0 and 1.3μm which the proposed Thematic Mapper does not at present include. The cross correlation tables thus derived are shown in Tables 2.4-12 thru 2.4-17. Specifications for the data sets used are given in the Table captions.

The correlation of T.M. band .52-.60μm and MSDS band .57-.63μm for the agricultural crops given in the tables ranged between .93 and .99. This tends to indicate that the use of the MSDS band .57-.63μm can represent the T.M. band reasonably well, even though the MSDS band includes part of the slope between the green peak reflectance and the chlorophyll absorption band in the red.

Concern has also been raised that T.M. bands .74-.80μm and .80-.91μm are highly correlated or more strongly stated - entirely redundant. The results of the correlation study support other studies showing that the channels are highly correlated. The correlation of the two channels ranged between .981 and .998. The plot shown in Figure 2.4-15 illustrates the high degree of correlation in the simulated channels using the MSDS data for the entire 8/15 North Dakota flightline.

Table 2.4-11

Results of Classifications Obtained for the Spectral Band Parameter Using Simulated 30 Meter Data.

Site - Date	Feature Set (Channels)			
	2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,6,7,8	2,3,4,5
Overall Train Performance (per cent correct)				
Kansas - 7/6	98.8	99.3	99.0	90.5
N. Dakota - 8/15	97.5	98.1	97.9	91.4
Overall Test Performance (per cent correct)				
Kansas - 7/6	88.8	88.9	88.2	81.3
N. Dakota - 8/15	90.7	91.2	91.4	86.8
RMS Error of Informational Class Proportion Estimates of Flightline (per cent)				
Kansas - 7/6	2.1	2.4	2.2	3.4
N. Dakota - 8/15	1.4	1.0	0.9	3.0
Average RMS Error of Informational Class Proportion Estimates of Twelve Sections (per cent)				
Kansas - 7/6	4.5	4.5	4.4	5.7
N. Dakota - 8/15	4.7	4.4	4.2	5.6

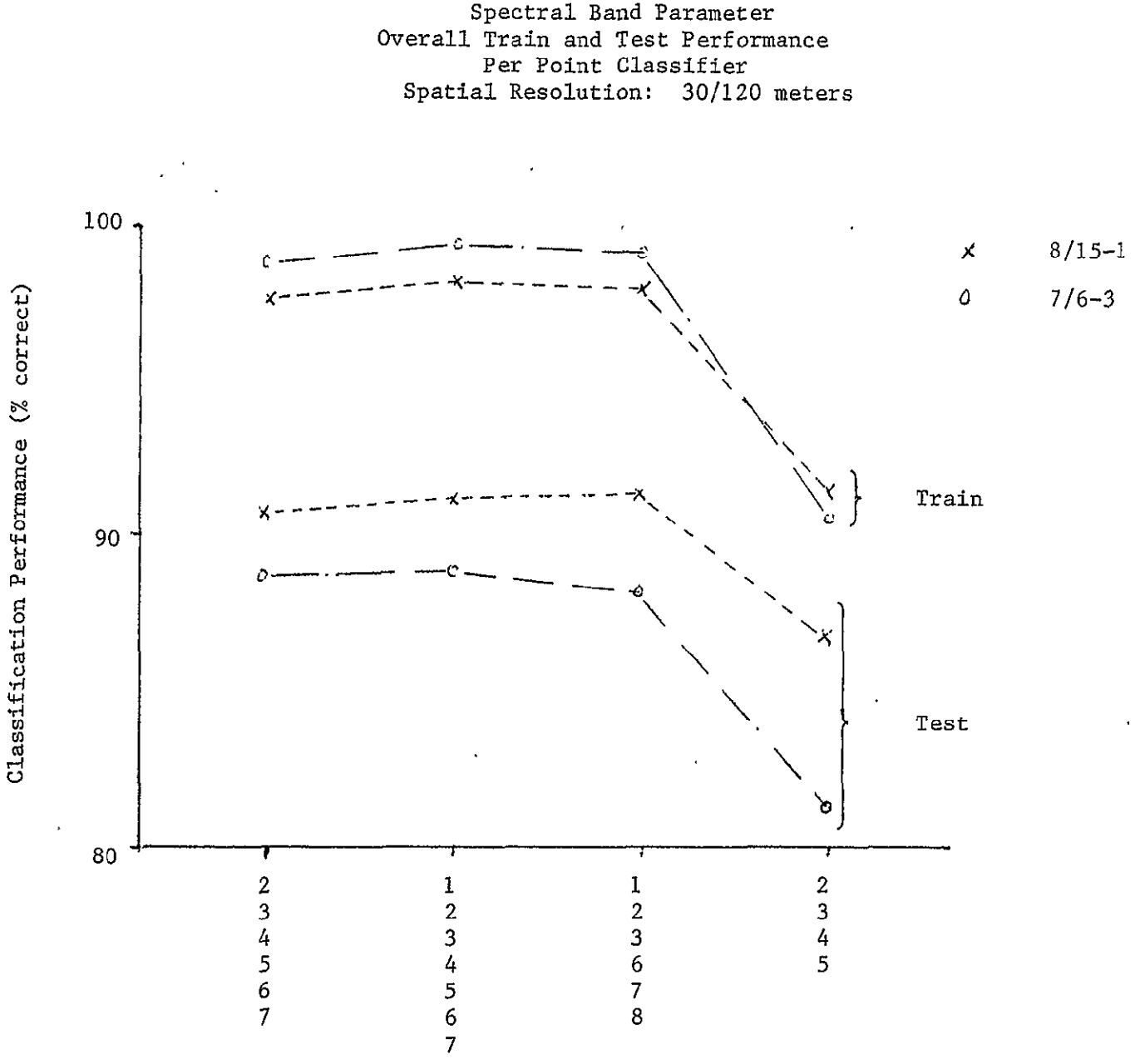


Figure 2.4-13 Feature Set (T.M. Channels)

Spectral Band Parameter
RMS Error of Proportion Estimates of Flight Line and Sections
Per Point Classifier
Spatial Resolution: 30/120 meters

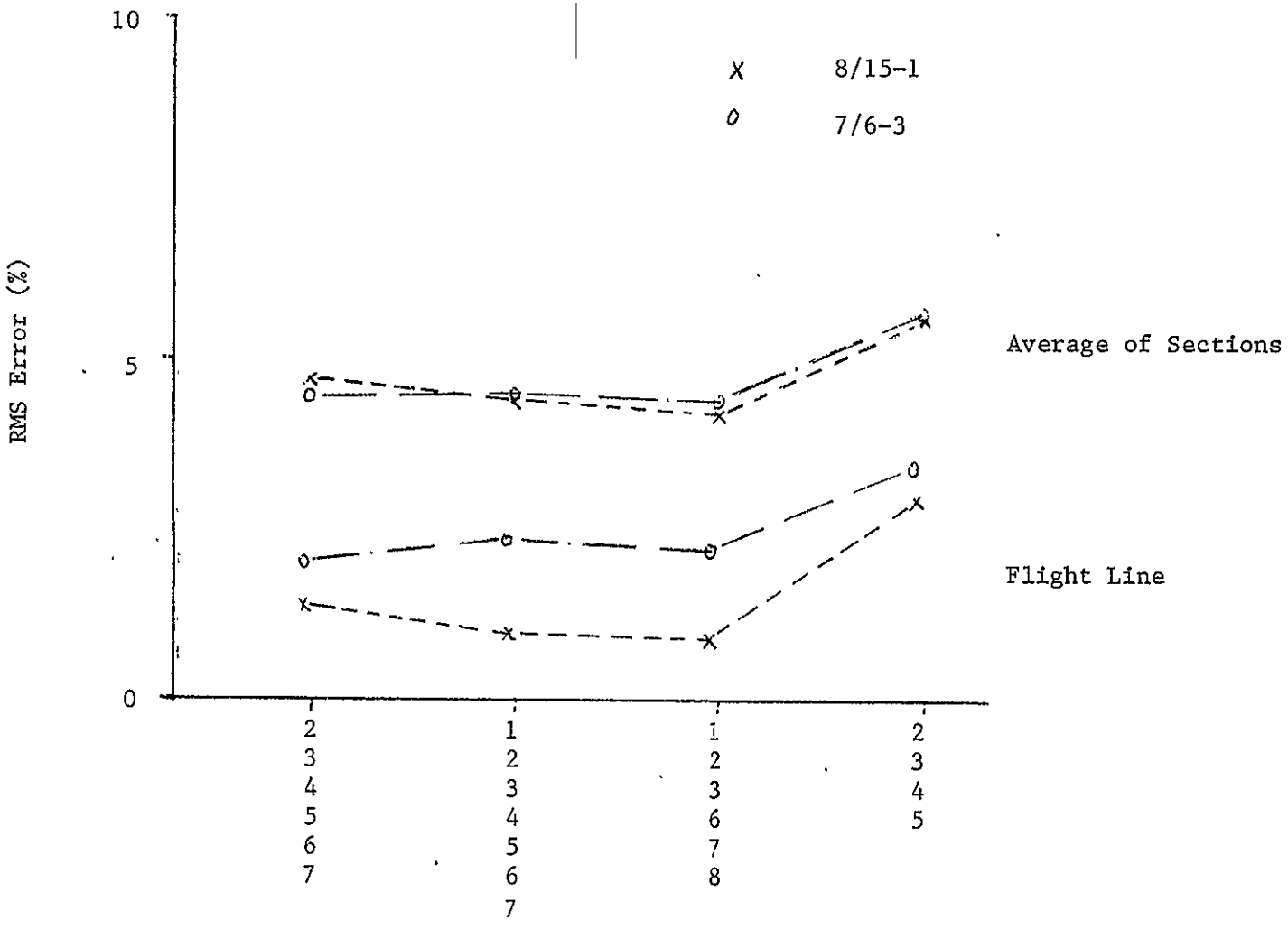


Figure 2.4-14 Feature Set (T.M. Channels)

Table 2.4-12

Correlation of Selected Wavelength Bands Using Spectrometer Data

Location: Purdue University Agronomy Farm, West Lafayette, Indiana
Instrument: Exotech 20C
Dates: July 6 - August 29, 1972; July 6 - October 5, 1973;
July 16 - August 15, 1974

Crop	No. Observations
Corn	353
Soybeans	105
Bare Soil	66

CORRELATION MATRIX

SPECTRAL BAND	0.52 1.60	0.57 0.63	0.63 0.69	0.74 0.80	0.80 0.91	0.98 1.08	1.09 1.19	1.20 1.30	1.55 1.75
0.52 0.60	1.000								
0.57 0.63	0.993	1.000							
0.63 0.69	0.963	0.987	1.000						
0.74 0.80	-0.020	-0.064	-0.120	1.000					
0.80 0.91	-0.038	-0.080	-0.136	0.981	1.000				
0.98 1.08	0.015	-0.022	-0.069	0.958	0.941	1.000			
1.09 1.19	0.040	0.010	-0.026	0.923	0.904	0.976	1.000		
1.20 1.30	0.092	0.084	0.063	0.848	0.841	0.904	0.936	1.000	
1.55 1.75	0.252	0.301	0.348	0.292	0.281	0.423	0.511	0.719	1.000

Table 2.4-13

Correlation of Selected Wavelength Bands Using Spectrometer Data

Location: Agricultural Research Farm, Garden City, Kansas
Instrument: Exotech 20C
Dates: October 18 - November 5, 1974

Crop	No. Observations	Crop	No. Observations
Wheat	45	Soybeans	6
Grain Sorghum	26	Alfalfa	5
Sugar Beets	14	Barley	3
Bare Soil	3	Rye	3
Corn	9		

CORRELATION MATRIX

SPECTRAL BAND	0.52 0.60	0.57 0.63	0.63 0.69	0.74 0.80	0.80 0.91	0.98 1.08	1.09 1.19	1.20 1.30	1.55 1.75
0.52 0.60	1.000								
0.57 0.63	0.947	1.000							
0.63 0.69	0.807	0.950	1.000						
0.74 0.80	0.017	-0.182	-0.350	1.000					
0.80 0.91	-0.000	-0.177	-0.319	0.990	1.000				
0.98 1.08	0.029	-0.096	-0.194	0.936	0.973	1.000			
1.09 1.19	0.115	0.023	-0.055	0.871	0.920	0.982	1.000		
1.20 1.30	0.311	0.304	0.276	0.641	0.704	0.833	0.918	1.000	
1.55 1.75	0.567	0.753	0.841	-0.421	-0.392	-0.233	-0.061	0.328	1.000

Table 2.4-14

Correlation of Selected Wavelength Bands Using Spectrometer Data

Location: Intensive Test Site, Finney County, Kansas
Instrument: FSS/S191H
Date: November 5, 1974

Crop	No. Observations	Crop	No. Observations
Wheat	1073	Pasture	10
Alfalfa	76	Grain Sorghum	102
Corn	248	Fallow	152

CORRELATION MATRIX

SPECTRAL BAND	0.45 0.52	0.52 0.60	0.57 0.63	0.63 0.69	0.74 0.80	0.80 0.91	0.98 1.08	1.09 1.19	1.20 1.30	1.55 1.75
0.45 0.52	1.000									
0.52 0.60	0.965	1.000								
0.57 0.63	0.956	0.984	1.000							
0.63 0.69	0.909	0.946	0.985	1.000						
0.74 0.80	-0.028	0.155	0.051	0.041	1.000					
0.80 0.91	-0.078	0.106	0.011	0.014	0.994	1.000				
0.98 1.08	-0.019	0.166	0.091	0.113	0.962	0.977	1.000			
1.09 1.19	-0.014	0.142	0.065	0.093	0.920	0.934	0.958	1.000		
1.20 1.30	0.074	0.209	0.147	0.184	0.840	0.854	0.905	0.949	1.000	
1.55 1.75	0.782	0.816	0.851	0.878	0.075	0.065	0.171	0.136	0.255	1.000

Table 2.4-15

Correlation of Selected Wavelength Bands Using Spectrometer Data

Location: Agriculture Research Farm, Williston, North Dakota
Instrument: Exotech 20C
Dates: June 2 - July 9, 1975

Crop	No. Observations	Crop	No. Observations
Oats	4	Fallow	2
Barley	5	Grass	3
Purum	4	Wheat	55

CORRELATION MATRIX

SPECTRAL BAND	0.52 0.60	0.57 0.63	0.63 0.69	0.74 0.80	0.80 0.91	0.98 1.08	1.09 1.19	1.20 1.30	1.55 1.75
0.52 0.60	1.000								
0.57 0.63	0.969	1.000							
0.63 0.69	0.929	0.985	1.000						
0.74 0.80	-0.707	-0.786	-0.809	1.000					
0.80 0.91	-0.706	-0.785	-0.804	0.998	1.000				
0.98 1.08	-0.703	-0.777	-0.799	0.998	0.997	1.000			
1.09 1.19	-0.655	-0.734	-0.742	0.982	0.988	0.985	1.000		
1.20 1.30	-0.635	-0.705	-0.723	0.977	0.977	0.985	0.984	1.000	
1.55 1.75	0.738	0.817	0.849	-0.670	-0.661	-0.639	-0.564	-0.511	1.000

Table 2.4-16

Correlation of Selected Wavelength Bands Using Spectrometer Data

Location: Agricultural Research Farm, Garden City, Kansas
Instrument: FSAS/Interferometer
Dates: October 19 - November 25, 1974; October 1, 1975

Crop	No. Observations	Crop	No. Observations
Wheat	22	Soybeans	3
Wheat Stubble	1	Alfalfa	3
Grain Sorghum	8	Surgar Beets	30
Corn	4	Grass	1

CORRELATION MATRIX

SPECTRAL BAND	0.45 0.52	0.52 0.60	0.57 0.63	0.63 0.69	0.74 0.80	0.80 0.91	0.98 1.08	1.09 1.19	1.20 1.30	1.55 1.75
0.45 0.52	1.000									
0.52 0.60	0.937	1.000								
0.57 0.63	0.939	0.974	1.000							
0.63 0.69	0.912	0.925	0.985	1.000						
0.74 0.80	-0.708	-0.596	-0.727	-0.792	1.000					
0.80 0.91	-0.710	-0.594	-0.717	-0.775	0.997	1.000				
0.98 1.08	-0.668	-0.537	-0.652	-0.707	0.975	0.988	1.000			
1.09 1.19	-0.596	-0.437	-0.556	-0.616	0.934	0.953	0.984	1.000		
1.20 1.30	-0.301	-0.109	-0.218	-0.281	0.713	0.748	0.833	0.906	1.000	
1.55 1.75	0.866	0.877	0.932	0.942	-0.806	-0.788	-0.705	-0.597	-0.212	1.000

Table 2.4-17

Correlation of Selected Wavelength Bands Using Spectrometer Data

Location: Agricultural Research Farm, Garden City, Kansas
Instrument: Exotech 20C
Dates: October 18 - November 5, 1974

Crop	No. Observations	Crop	No. Observations
Wheat	33	Corn	6
Grain Sorghum	17	Soybeans	3
Sugar Beets	11	Alfalfa	1
Bare Soil	3		

CORRELATION MATRIX

SPECTRAL BAND	0.52 0.60	0.57 0.63	0.63 0.69	0.74 0.80	0.80 0.91	0.98 1.08	1.09 1.19	1.20 1.30	1.55 1.75	10.40 12.50
0.52 0.60	1.000									
0.57 0.63	0.928	1.000								
0.63 0.69	0.753	0.940	1.000							
0.74 0.80	-0.137	-0.371	-0.533	1.000						
0.80 0.91	-0.190	-0.392	-0.515	0.985	1.000					
0.98 1.08	-0.185	-0.314	-0.374	0.907	0.962	1.000				
1.09 1.19	-0.097	-0.177	-0.207	0.823	0.895	0.978	1.000			
1.20 1.30	0.156	0.197	0.232	0.504	0.596	0.769	0.881	1.000		
1.55 1.75	0.575	0.807	0.907	-0.629	-0.610	-0.456	-0.276	0.196	1.000	
10.40 12.50	0.078	0.207	0.284	-0.207	-0.199	-0.144	-0.082	0.052	0.312	1.000

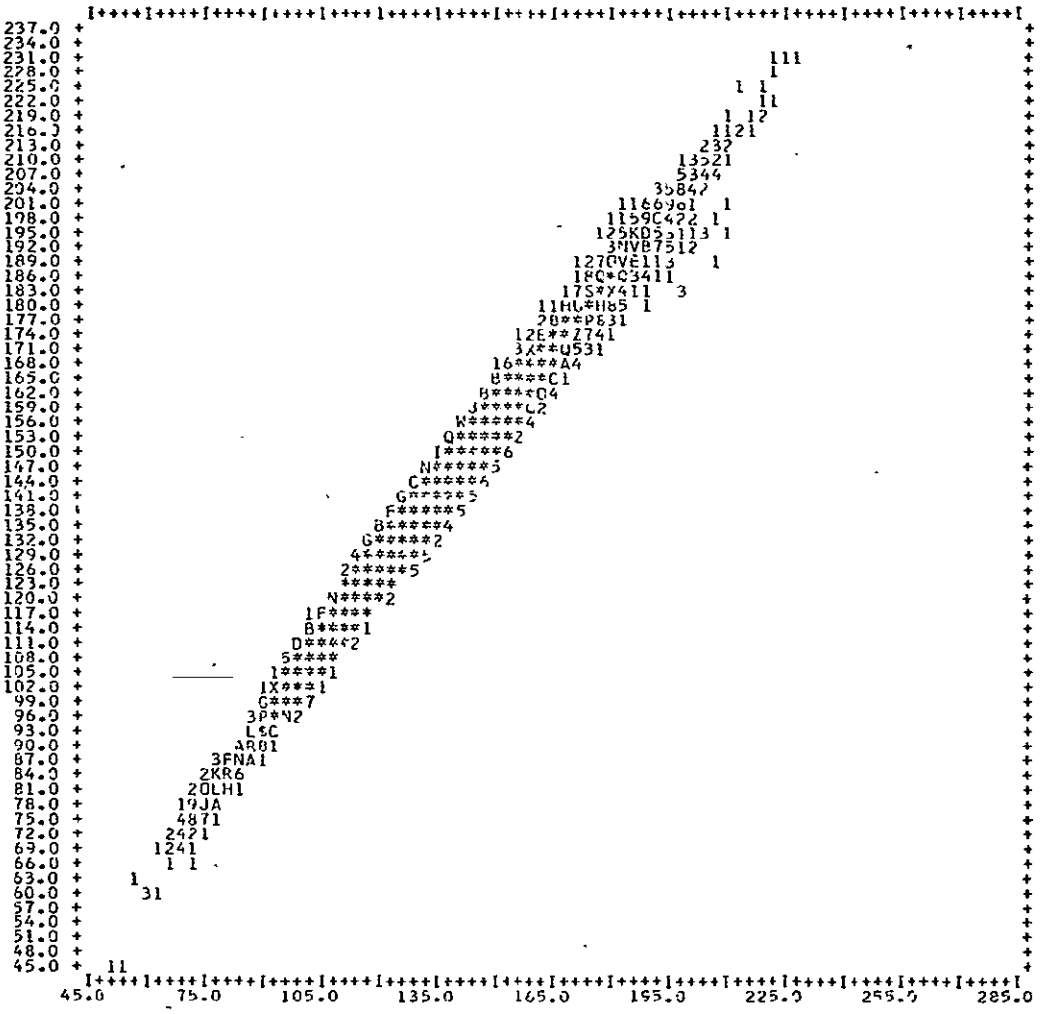


Figure 2.4-15 Scatter diagram showing correlation between channels 4 and 5.

Another item of concern about the proposed Thematic Mapper bands is the lack of any bands in the 1.0-1.3μm range. Information in this range may not be available in the other bands. The results for the observations analysed support the above concern. The channels in the 1.0-1.3μm range were correlated the most with Thematic Mapper channels 4 and 5. However, the correlation of the 1.09-1.19μm band and Thematic Mapper bands 4 and 5 ranges between .87 and .99. More significantly the correlation of the 1.2-1.3μm band and Thematic Mapper channels 4 and 5 ranges between .50 and .98. The results suggest that useful information may be available in the 1.0-1.3μm range.

In general, Tables 2.4-12 thru 2.4-17 indicate that Thematic Mapper channel 6, the middle IR channel, and Thematic Mapper channel 7, the thermal channel, are not very correlated with any of the other Thematic Mapper channels. The visible channels tend to be correlated and Thematic Mapper channels 4 and 5 are highly correlated. There may be good reasons to move one of the .74-.80μm or .80-.91μm bands into the 1.0-1.3μm range.

Classifier Parameter - Two different classifiers were compared - the standard maximum likelihood pixel classifier and a spectral-spatial classifier ECHO[4]. The simulated data over the North Dakota 8/15 flightline and the Kansas 7/6 flightline were used in the analysis. The training statistics for both classifiers were identical and were obtained as described in the spatial resolution parameter discussion.

The classifiers were compared across the four simulated spatial resolutions for the North Dakota flightline and across four noise levels for the Kansas flightline. The four criteria as described before were used to evaluate the classifiers across the spatial resolutions. The train and test performances were used across the noise levels. The results are illustrated in Figures 2.4-16 thru 2.4-19.

There was a slight but consistant increase in the training and test performances for the ECHO classifier over the per point classifier at the smaller IFOV's (Figure 2.4-16). This is consistent with the theory behind the ECHO classifier. Better classification accuracy should be obtainable as the number of pixels per object (field) increases. The differences between the classifiers are so small, however, that they may not be significant for this particular case.

A very noticeable difference was observed, however, when comparing the classifiers across the noise levels (Figures 2.4-18 and 2.4-19). The spatial nature of the ECHO classifier was able to provide enough information to help compensate for the added noise in the data. The difference between the two classifiers became greater as the noise level increased.

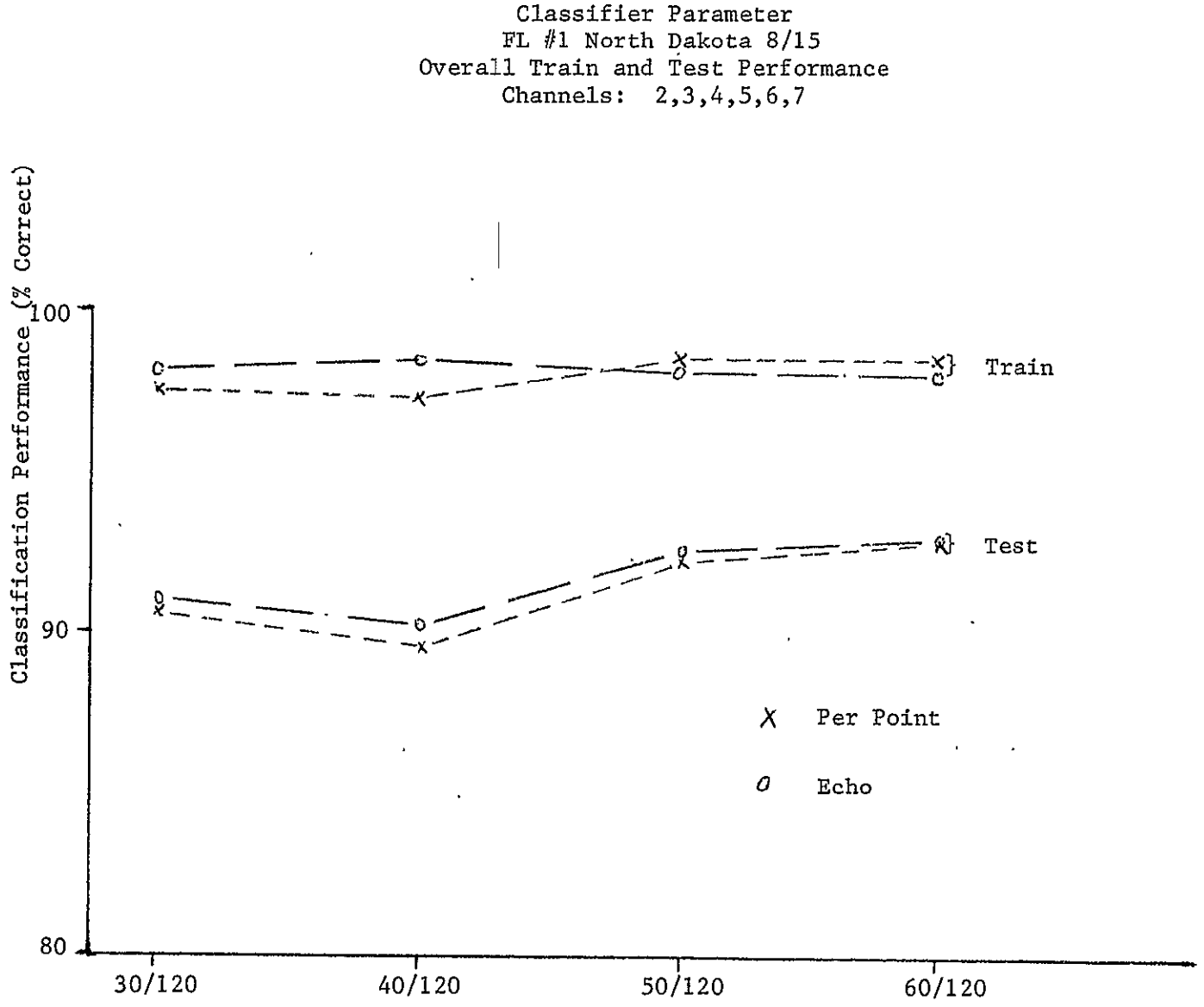


Figure 2.4-16 Reflective/Emissive Resolution (m)

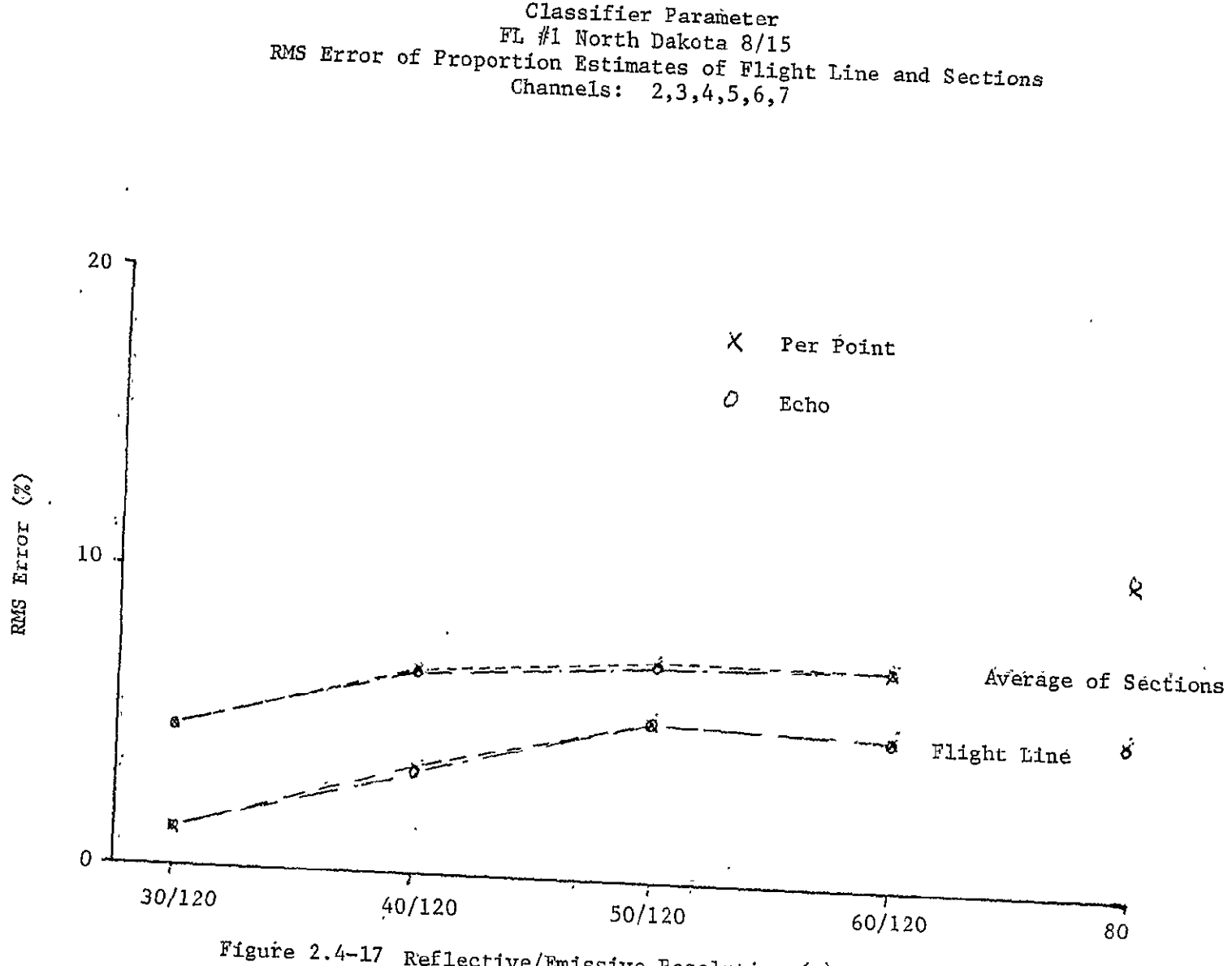


Figure 2.4-17 Reflective/Emissive Resolution (m)

Classifier Parameter
FL #3 Kansas 7/6
Overall Train and Test Performance
Channels: 2,3,4,5,6,7
Resolution: 30/120 meters
Noise for 0 added set .004-.023 NEΔP, NEΔT

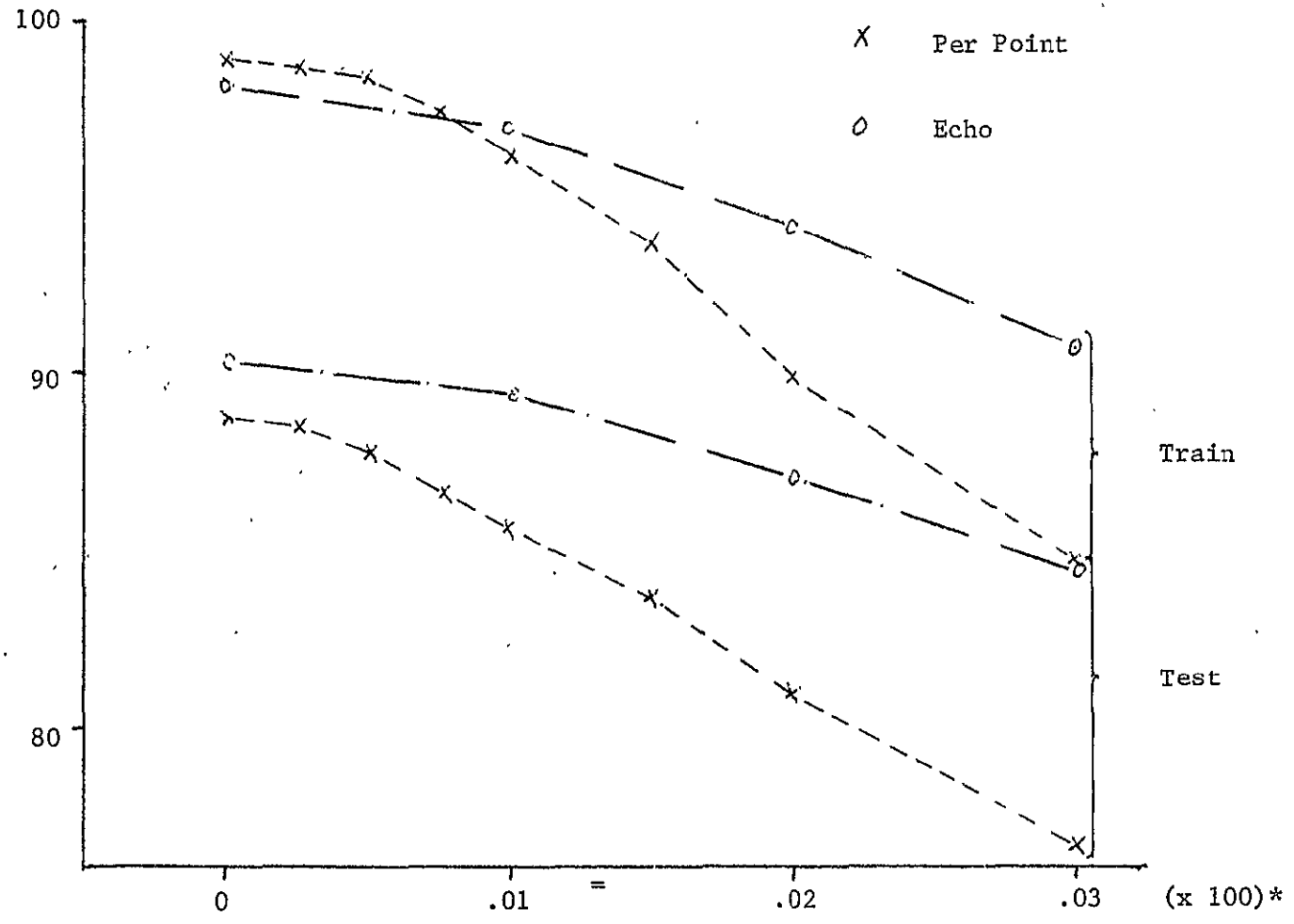
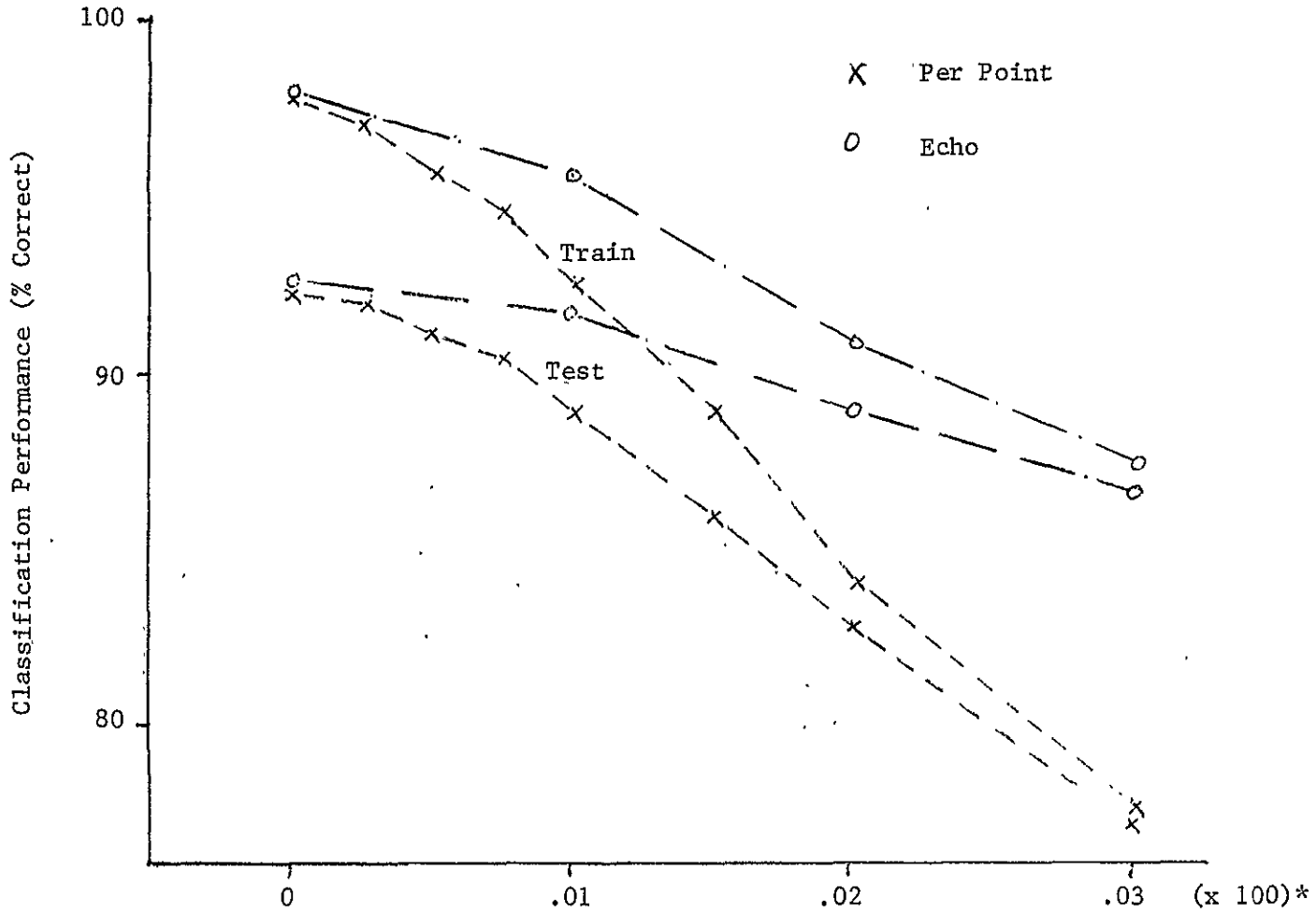


Figure 2.4-18 Noise Added (NEΔP or *NEΔT)

Classifier Parameter
FL #3 Kansas 7/6
Overall Train and Test Performance
Channels: 2,3,4,5,6,7
Resolution: 40/120 meters



The results indicate that as spectral-spatial classifier is an improvement over the per point classifier. More research needs to be done with the ECHO classifier, however, to determine the affects of the many parameters that can be changed in the ECHO classifier. It is possible that the optimum classifications using the ECHO classifier was not obtained because of the lack of knowledge concerning the ECHO parameters.

SUMMARY AND CONCLUSIONS

In the introduction, the five sets of parameters which influence the ability to extract information from multispectral data were listed and it was pointed out that the problem of properly selecting scanner parameter values amounts to searching the five dimensional parameter space thus defined relative to the desired index performance. This study was structured, within the constraints imposed by the data sets available, to search a portion of this five dimensional space. The effect of at least some variation in all but the fourth parameter class was tested. Significant features of the study were as follows:

1. The index of performance used encompassed both identification accuracy and mensuration accuracy.

2. Data from two times of the year was used.

3. Data from two quite different parts of the U.S. Wheat Belt was used. Even so only a small part of world agriculture and world vegetation was sampled.

4. The impact of the affect of a human analyst was allowed in the study in that three different analysts, using slightly different analysis techniques, were used. As would be desired, there is no indication that this affected the results.

Both training sample accuracy and test sample accuracy were used to evaluate the various tests, the former because it tends to minimize the impact of variations in the scene. However, in this study it was possible to achieve such high training accuracies that the residual variations in performance shown by training sample accuracy tended to be random and the test sample accuracy appears to provide the more reliable indicator of the impact of the various parameters on identification accuracy. The two RMS proportion estimation error indicators were devised to provide the indication of combined identification and mensuration performance. Of these two, the RMS proporation estimation error for the entire flightline appears to provide the more sensitive indication to the impact of changing the various parameters of interest as compared to the average RMS error by sections.

The major conclusions from the study are as follows:

1. There was a very small but consistent increase in identification accuracy as the IFOV was enlarged. This is presumed to stem primarily from the small increase in signal to noise ratio with increase in IFOV.

2. There was a more significant decrease in the mensuration accuracy as the IFOV was enlarged.

3. The noise parameter study proved somewhat inconclusive due to the greater amount of noise present in the original MSDS data than desired. For example, viewing Figure 2.4-11 moving from right to left, it is seen that the classification performance continues to improve as the amount of noise added is decreased until the point is reached where the noise added approximately equals that already initially present due to the MSDS operation. Thus, it is difficult to say for what signal to noise ratio a point of diminishing return would have been reached had the initial noise not been present.

4. The result of the spectral band classification studies may also be clouded by the noise originally present in the MSDS data. The relative amount of that change in performance due to using different combinations of the .45-.52μm, .74-.80μm, .80-.91μm and .74-.91μm bands is slight but there appears to be a slight preference for the .45-.52μm band. The performance improvement of the Thematic Mapper channels over those approximating Landsat I/II is clear however.

5. Using spectrometer data it was verified that the .74-.80μm and .80-.91μm bands are highly correlated.

6. Correlation studies also showed that the range from 1.0-1.3μm is likely to be an important area in discriminating between earth surface features.

Although much has been learned in this study about the selection of parameters for the Thematic Mapper, it is clear that this problem cannot be now regarded as entirely solved. Further studies of this and other types are needed to develop a convincing set of facts regarding scanner system parameters selection. This study also illustrates very clearly the value of both field gathered and airborne multispectral data in continuing research efforts.

REFERENCES

1. Harnage, J. L. and Landgrebe, D. A., June 1975
Landsat-D Thematic Mapper Technical Working Group - Final Report
NASA Johnson Space Center Report No. JSC-09797, Houston, Texas

2. Zactzeff, Eugene M., C. Korb, and Charles Wilson, July 1971.
MSDS: An Experimental 24-Channel Multispectral Scanner System
IEEE Transactions of Geoscience Electronics, Vol. GE-9, No. 3

3. BMD Biomedical Computer Programs, 1970
University of California Publications in Automatic Computation No. 2
University of California Press, Berkeley and Los Angeles, CA

4. Kettig, R. L. and Landgrebe, D. A., January 1976
Classification of Multispectral Image Data by Extraction and
Classification of Homogeneous Objects
IEEE Transactions of Geoscience Electronics, Vol. GE-14, No. 1,
pp. 19-26

APPENDIX A

Algorithms Used in Thematic Mapper Simulation Study

A. Pixel size in scan line direction (See Figure 1)

$$L_i = A * (\tan \Theta_1 - \tan \Theta_2) \qquad \text{(eq. 1)}$$

$$\text{where } \Theta_1 = \alpha - \Delta * (i-1)$$

L_i - Pixel length in scan line direction, element i

A - Scanner altitude

Θ - as shown in Figure 1

α - maximum scan angle off nadir

Δ - scanner instanteous field of view

B. Pixel size in down track direction

$$H = \frac{A}{\cos \Theta_i}$$

$$\frac{W_i}{2} = H * \tan \left(\frac{\Delta}{2}\right) \qquad \text{(eq. 2)}$$

W_i - Pixel width in down track direction, element i

C. Weighed averaging of pixels in the down track direction. (See Figure 2).

High resolution pixels are averaged to form each low resolution pixel. In the down track direction unweighted averaging is used except when a high resolution pixel occurs only partially within the low resolution swath. Response values (R) are computed for each high resolution image column (i) corresponding to a low resolution swath. The response or count for the ground area L_i long (cross track high resolution pixel size) and down track dimension equal to one low resolution swath width is the proportion (P) of each input pixel of column i falling within the swath multiplied by its response and divided by the sum of the proportions.

$$R_i = \left(\sum_{j=1}^{J} R_{ij} * P_{ij}\right) / \sum_{j=1}^{J} P_{ij} \qquad \text{(eq. 3)}$$

MSDS Scanline Ground Coverage

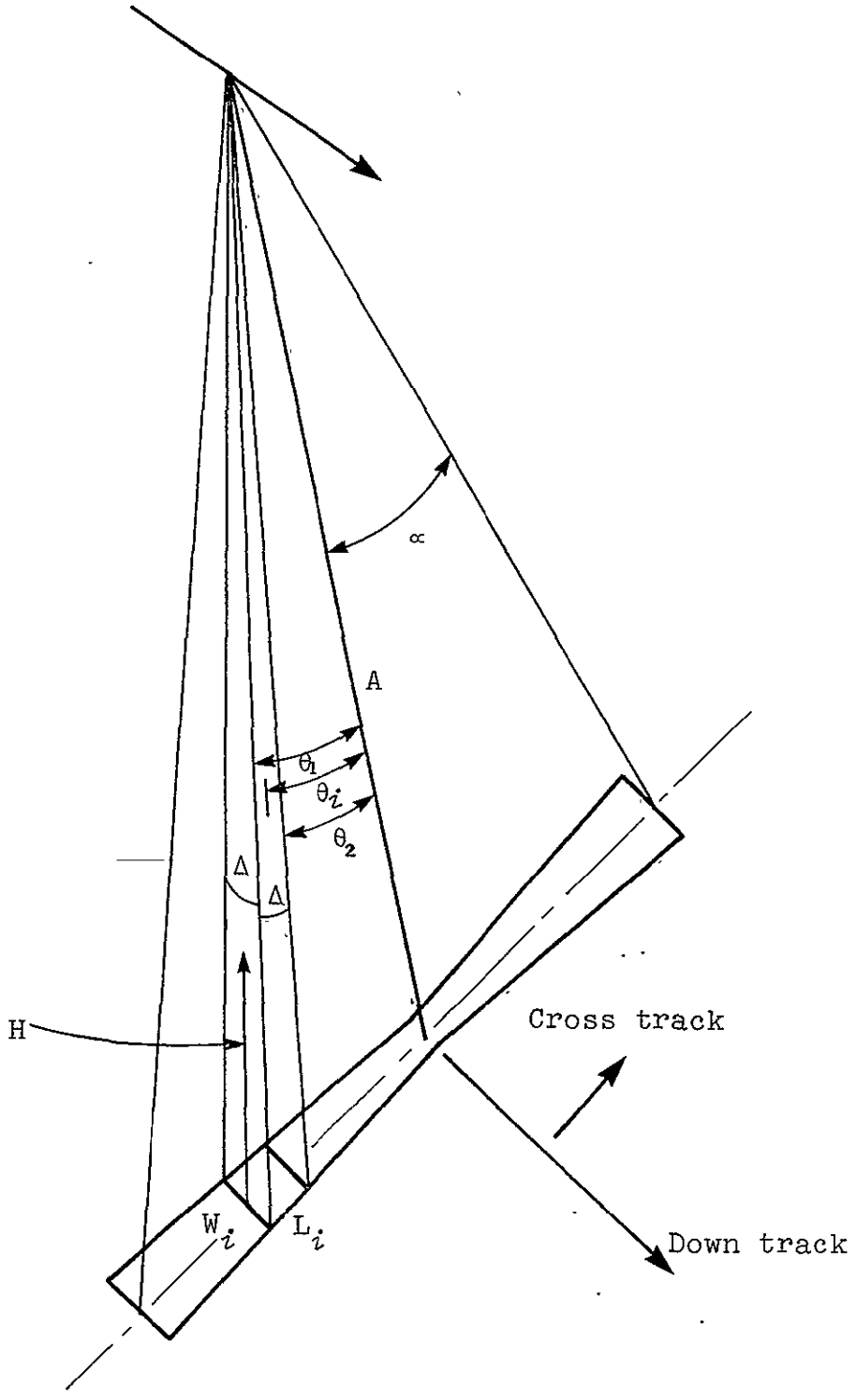


Figure 1. The MSDS ground coverage for one scan line

MSDS Sample Size

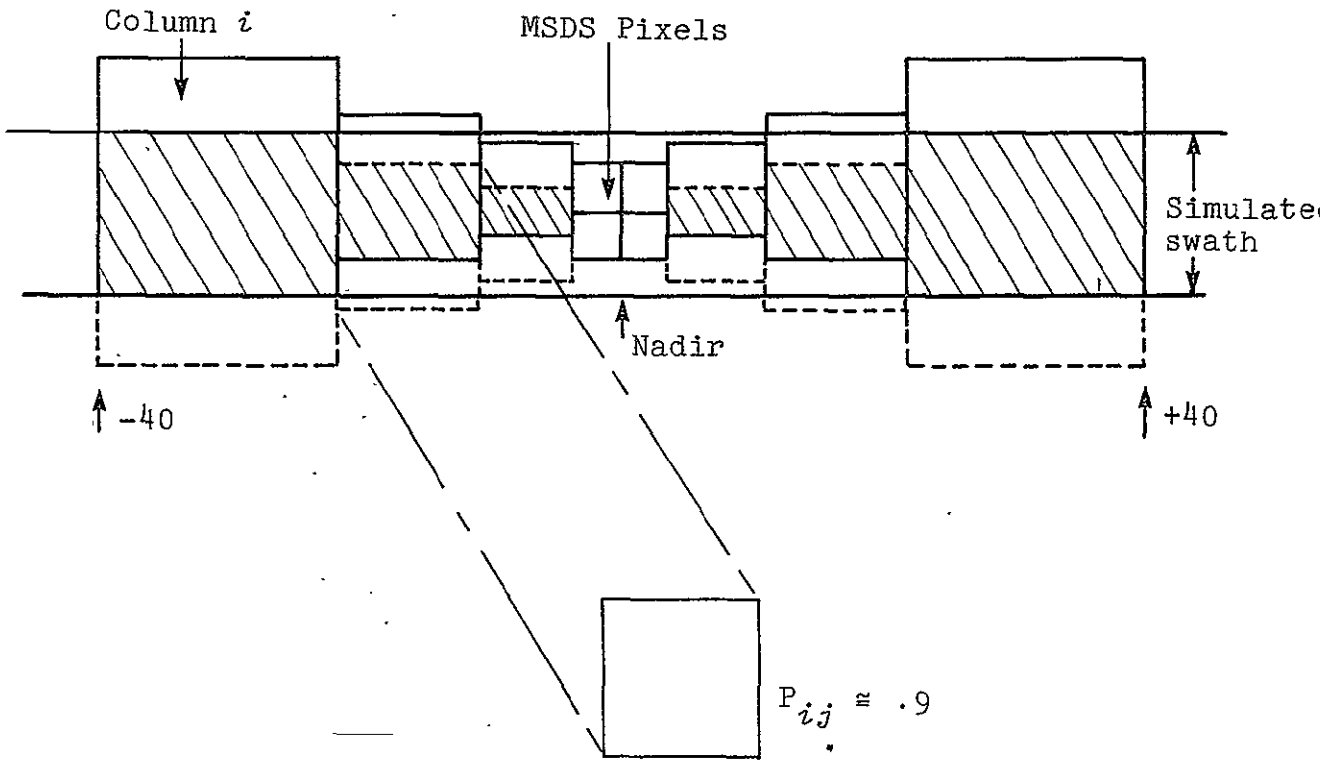


Figure 2a. MSDS samples overlap for off nadir look angles

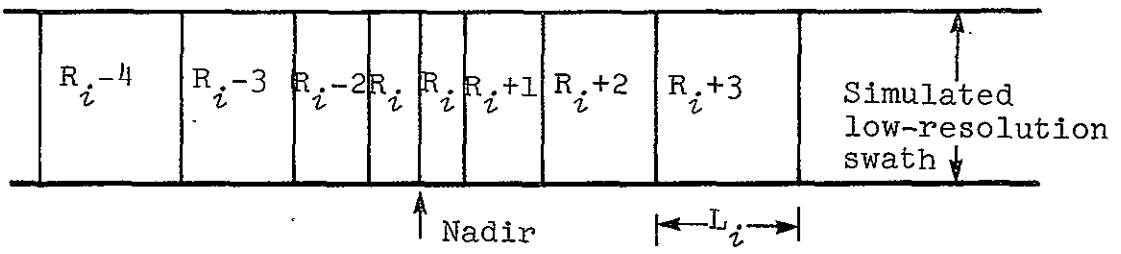


Figure 2b. MSDS samples represent increased ground coverage at off nadir look angles

R - response from ground scene area L_i by simulated swath width

J - high resolution scan lines included in simulated width

j - high resolution line

i - high resolution column or sample

P - proportion included in simulated swath

Figure 2b illustrated the ground areas corresponding to the R values.

D. Weighted averaging of pixels in the cross track direction.

For data simulation purposes the cross track simulated sample is a Gaussian weighted average of high resolution pixels, with the half amplitude point of the Gaussian distribution function a distance of one half IFOV from the distribution midpoint. Only R values from equation 3 above are considered for the cross track weighting.

The Gaussian distribution function has the form:

$$G(\mu,\sigma,x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-5(\frac{x-\mu}{\sigma})}$$

G - weight for averaging

μ - low resolution pixel center

x - high resolution pixel distance from simulated pixel center

σ - constant for a simulation resolution S

S - simulation resolution

Sitting μ = o and x = one half ground IFOV = S/2 will enable solution for the constant σ. The .1 amplitude point (x(.1) the distance from the mean or low resolution pixel center) can then be calculated.

Low resolution pixel centers are equally spaced across the input image width with spacing S. This assumes all low resolution samples are at satellite scanner nadir.

The data value of each low resolution element for a given simulated swath is then

$$R'_k = \frac{\Sigma(G(\mu,\sigma,x_i)*R_i*L_i)}{\Sigma G(\mu,\sigma,x_i)*L_i}$$

R' - low resolution pixel value or count

k - simulated pixel number

The summations are made over columns (i) for which G is greater than .1. Figure 3 illustrates the weighting function over the high resolution average columns.

Reflectance Calibration

E. C_{lb} - Mean over the run of low cal black body

C_l - Mean over the run of lamp cal

R_e - Lamp equivalent reflectance

x_i - MSDS count at sample i

R_i - Reflectance value at sample i

$$R_i = (x_i - C_{lb}) \cdot \frac{R_e}{(C_l - C_{lb})}$$

$$= \frac{R_e}{(C_l - C_{lb})} \cdot x_i - \frac{R_e \cdot C_{lb}}{C_l - C_{lb}}$$

$$\therefore \quad R_i = Ax_i + B \qquad \text{(eq. 1)}$$

$$\text{for } A = \frac{R_e}{C_l - C_{lb}} \quad \text{and } B = \frac{-R_e \cdot C_{lb}}{C - C_{lb}}$$

Note: x_i is the count of sample i of the degraded resolution image: Resolution degradation was handled immediately preceding.

Simulated Reflectance Range Transformation

F. R_s - Simulated full scale saturation reflectance

x_i - Simulated count at sample i

$$x_i = 255 * \frac{R_i}{R_s}$$

Simulation Weighting Function
and
High Resolution Pixel Column Averages

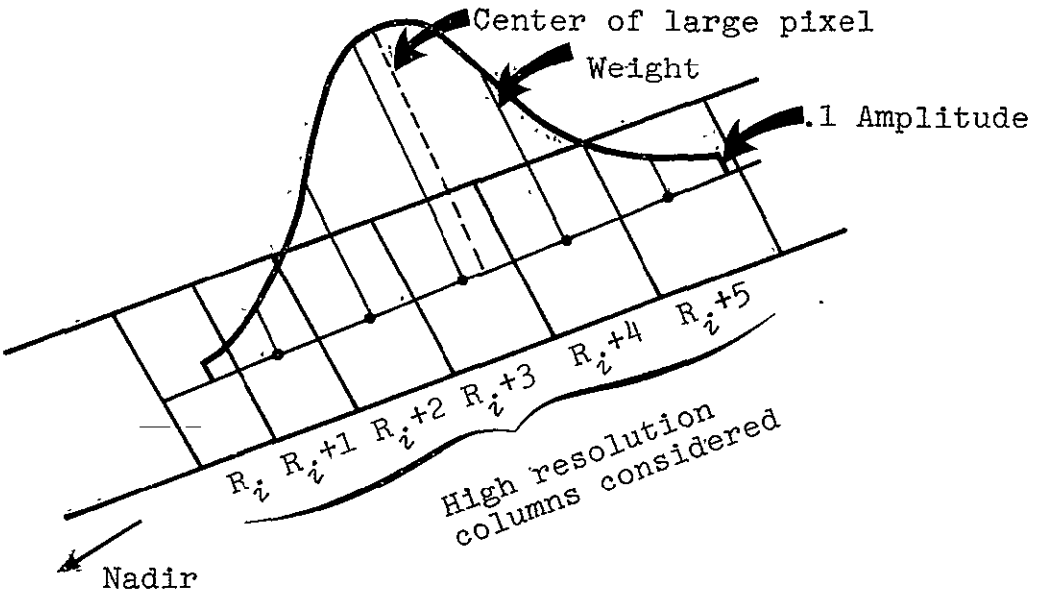


Figure 3. High resolution pixels are averaged in the down track direction to form columns and columns are averaged by the weighting function

$$= \frac{255}{R_s} * R_i$$

$$x_i = \frac{255}{R_s} (Ax_i + B)$$

$$= \left(\frac{255}{R_s} * \frac{R_e}{(C_1 - C_{1b})}\right) x_i + \left(\frac{255}{R_s}\right)\left(\frac{-R_{e*}C_{1b}}{C_1 - C_{1b}}\right)$$

$$x_i = \frac{255}{R_s} \frac{R_e}{(C_1 - C_{1b})} \cdot \left(x_i - C_{1b}\right) \qquad \text{(eq. 2)}$$

To x_i noise must be added and we have a "SUPER PIXEL"

The program will allow for a noise level addition parameter to be input for each run. The value will be a floating point number equal to the noise standard deviation.

G. Addition of thermal channels

MSDS bands 10.1 - 11.0 micrometers and 11.0 - 12.0 micrometers were combined by performing a weighted average. Weights were based on the proportion of radiant energy produced by a 300 ^{o}K blackbody at the band mid-points. This method requires that the raw data points be calibrated into radiance units prior to combination. Radiance calibration is performed as follows:

$$R_i = A_0 + A_1 * x_i$$

$$A_1 = \frac{R_H - R_L}{M_H - M_L}$$

$$A_0 = R_L - M_L * A_1$$

where:

R_i = radiance of element i

x_i = relative response of element i

R_H = radiance of high blackbody cal.

R_L = radiance of low blackbody cal.

M_H = mean relative response of high cal.

M_L = mean relative response of low cal.

M_H and M_L are averaged over the full flightline. R_H and R_L are given by the Planck blackbody radiance function.

$$R_H = P(T_H)$$

$$P(T) = \frac{1}{\lambda_{b_i}} \int_{\lambda_1}^{\lambda_2} \frac{C_1 d\lambda}{\lambda^5 \, e^{(C_2/\lambda T)} - 1}$$

$C_1 = 1.1909 \cdot 10^4$ watts-$\mu m^4/cm^2$-ster

$C_2 = 1.4388 \cdot 10^4$ μm ^oK

λ_1, λ_2 = band wavelength limits

T_H, T_L - Thermal temperature of high, low cals.

λ_{bi} = Bandwidth of λ_1-λ_2

Weighting coefficients for computing the radiance of the combination band is given by the ratio of spectral radiant existance for the bands. The ratio is:

$$M_\lambda = \frac{\varepsilon C_1}{\lambda^5 (e^{C_2/\lambda T} - 1)}$$

$$\therefore \frac{M_{\lambda_o}}{M_{\lambda_1}} = \left(\frac{\lambda_1}{\lambda_0}\right)^5 \frac{\left(e^{C_2/\lambda_1 T} - 1\right)}{\left(e^{C_2/\lambda_0 T} - 1\right)}$$

λ_0, λ_1 = band wavelength midpoints

T = average scene temperature

The combination radiance is then:

$$R_{\lambda_o \lambda_1} = a R_{\lambda_o} + b R_{\alpha 1}$$

$$X = \frac{M_{\lambda_o}}{M_{\lambda_1}}$$

$$a = \frac{X}{X+1} \qquad b = \frac{1}{X+1}$$

H. Addition of reflective bands.

MSDS bands .76 - .80 micrometers and .82 - .87 micrometers were combined to simulate the .74 - .91 micrometers band. Band addition was achieved by unweighted averaging.

$$R_\lambda = \frac{R_{\lambda_o} + R_{\lambda_1}}{2}$$

R = reflectance factor for a pixel

I. White Gaussian independent noise samples were added to the data such that the Thematic Mapper noise specifications are approximated. To compute the level of noise to be added, the existing noise level was first estimated. To do this, the 16 low blackbody calibration samples of each scene line were processed through the system separately as a 16 sample wide image. The standard deviation of the calibration samples after processing were used to define the data noise level. The appropriate noise level was then added to approximate Thematic Mapper noise specifications.

APPENDIX B

2.4 Scanner System Modelling

A number of facets to the scanner system modelling problem were considered early in the contract year. Among these were statistical representation of scene classes, spectral window functions, scanner aperature effects, sampling noise influences and classification error models. Since the majority of resources in the project was spent on the thematic mapper simulation only limited consideration was made of the generalized system elements. Specifically the classification error model was pursued in some detail in an attempt to advance the state of the art in the multiclass, multichannel classification error prediction problem. The following is a brief review of this work.

Probability of Misclassification for an Optimum Bayes Classifier Introduction

The evaluation of the performance of various classifiers through calculation of exact probability of misclassification has been under study by researchers for some time. Although empirical results have been obtained, the precise recognition rate for a Bayes classifier has not been evaluated for more than two normal classes with arbitrary covariance matrices (Anderson[1], Das Gupta[2], Fukunaga[3]). In evaluating the scanner performance, attention here has been focused on recognition rate at the both input and output of the systems. For this purpose, the probabilities of misclassifications of a multiple class Bayes classifier must be obtained. A Bayes classifier is a quadratic classification scheme which is optimum in the sense of providing correct probability of error. Much simpler mathematics is required for evaluation of a sub-optimum linear classifier.

Background

The discrimination problem, as it is known now, was introduced by R. A. Fisher as a problem in taxonomy. He used a linear discriminant function with coefficients chosen such that a defined cost function was minimized. Since then, many extensions to the problem have been proposed and solved. Bayes classification, however, has been almost entirely limited to two classes with multivariate normal distribution. As mentioned before, the source of the problem is the quadratic nature of discriminant functions.

Multiple Class Classification

For our purposes, error probability for a multiple class Bayes classifier is required. The problem is simplified somewhat by assuming that all distributions are multivariate normal and mean vector and covariance matrices are known. The absence of the latter assumption would immensely complicate the problem by introducing random matrices with Wishart distributions.

Approach

The basic approach is based on the following:

The quadratic discriminant functions are

$$\ell_i = (\underline{X}-\underline{M}_i)\Lambda_i^{-1}(\underline{X}-\underline{M}_i)-C_i \qquad i=1,2,.....M$$

where M_i and Λ_i are mean vector and covariance matrice of class i, C_i is a constant depending on prior probabilities of each class.

Now, choose class j if $\ell_i < \ell_i \ \forall j \neq i$

We defined the probability of error as:

Under M_1

$$\Pr(\varepsilon|M_1)_\infty = 1 - \Pr\{C|M_1\}_\infty = 1 - \Pr[\text{all } \ell_i > \ell_1 / \forall i \neq 1]$$

$$= 1 - \int_0^\infty f(\ell_1)\left[\int_{\ell_1}^\infty\int_{1_1}^\infty\int_{1_1} f(\ell_M, \ell_{M-1}.....\ell_2|\ell_1)\,d\ell n...d\ell_2\right]d\ell_1$$

where

$\Pr\{\varepsilon|M_1\}$ = Conditional probability of error

$\Pr\{C|M_1\}$ = Conditional probability of correct classification

As seen above, the quantities required are, the marginal and joint distributions of ℓ_i.

Assuming that the ovservation belongs to say, first class, i.e. $\underline{X} \tilde{} N(M_1, \Lambda_1)$ then ℓ_1 has a chi-squared distribution with N degrees of freedom (N=dimension of the observation vector). ℓ_2 and others have a more complicated distribution. Their moment generating function, however, can be found

$$E[e^{t\ell_2}] = \int e^{t[(X-M_2)^T\Lambda_2^{-1}(X-M_2)-C_2]} \frac{(X-M_1)^T\Lambda_1^{-1}(X-M_1)}{2} d\underline{X}$$

There is a single transformation that diagonalize both Λ_1^{-1} & Λ_2^{-1}. Carrying out the mathematics

$$M_{P_2}(t) = \prod_{i=1}^{N} \frac{e^{\frac{K_i t}{1-2di t}}}{1-2d_i t} \qquad |t| \leq \frac{1}{2di}$$

The density function of ℓ_2 is related to its m.g.f. in a known fashion.

Joint Distribution

Various methods have been tried with varying degree of success. The basic approach in all of them is based on calculating the m.g.f. of conditional random variables and by using the following chain formula:

$$f(Z_M Z_{M-1} \cdots Z_1) = f(Z_M | Z_{M-1}, Z_{M-2} \cdots Z_1)\ f(Z_{M-1} | Z_{M-2} \cdots Z_1)$$

$$\times\ f(Z_3 | Z_2, Z_1)\ f(Z_2 | Z_1)\ f(Z_1)$$

find the joint classify function

At the moment, we have focused on three classes. A method developed for this case will likely be extendable to more general case. It should be mentioned that, the limiting factor in this problem is not as much the number of classes as it is the dimension of observation vector.

The discriminant functions shown before, are hyperquadrics in M dimensions. To obtain $f(\ell_2 | \ell_1)$, ℓ_1 is fixed and any subsequent integration should bear in mind that the integrand is defined over this surface. In a special case, when N=3, the surface is a general ellipsoid.

The next case, $f(\ell_3 | \ell_2, \ell_1)$, both ℓ_2 and ℓ_1 are fixed, and the path of integration is on the intersection of these two hyperquadrics. If covariance matrices are equal, then each ℓ_i is a hypershpere on a sphere for N=3. Their intersection would be simple circles.

Conclusion

The direction that we are heading in at this point is to obtain $E[e^{t\ell 2} | \ell_1]$ and $E[e^{t\ell 3} | \ell_2, \ell_1]$ subject to above surface constraints. Beyond this, the inverse FFT would provide the desired conditional and joint density functions. It is unlikely that integrals are easily integrable, so numerical methods will likely be used throughout.

REFERENCES

1. Anderson, T. W., Bahadur, R. 1962.
Classifications into Two Multivariate Normal Distributions with Different Covariance Matrices
Annals of Math. Statistics, Vol. 33, pages 420-431.

2. Das Gupta, S. 1965.
Optimum Classification Rules for Classification into Two Multivariate Normal Populations.
Annals Math. Statistics, Vol. 36, pages 1174-1184.

3. Fukunaga, L., and Knile, T. March 1969.
Calculations of Bayes Recognition Error for Two Multivariate Gaussian Distributions.
IEEE Transactions on Computers.

2.5 Transfer of Computer Image Analysis Techniques (Remote Terminal)

Activities under this section of the work statement included continued support of the JSC-LARS Computer Facility remote terminal and the development of technology transfer concepts and materials. Work under this task was carried out in accordance with the milestone plan submitted during the first month of the contract period. For convenience, the discussion of these activities parallels the milestone plan which is summarized below.

	Activity	Time Period
I	Task Analysis	June-September, 1975
II	Training Program Design	July-October, 1975
III	Development of Priority Assignment	September-October, 1975
IV	Materials Development	October, 1975-May, 1976
V	Interim Training Programs	July, 1975-May, 1976
VI	Remote Terminal Support	June, 1975-May, 1976

Task Analysis Activities

On June 12, Ms. Shirley Davis and Ms. Tina Cary visited JSC and met with Mr. Donald Hay. The work statement and preliminary milestones for this task were reviewed at that time. On August 19 and 20, Dr. John Lindenlaub visited JSC to meet with Mr. Hay to discuss training needs, determine typical entry levels of trainees and survey presently available training materials (milestone I). A series of meetings between Dr. Lindenlau and Building 17 personnel was arranged by Mr. John Sargent. Most of the discussions were on a one-to-one basis with personnel working on specific projects. The exception was a group discussion organized by Mr. Tom Minter which centered around LACIE Analyst Interpreters (AI) and Data Processing Analyst (DPA) training.

The following observations were made as a result of the one-to-one conversations. The individuals talked to were for the most part working within a relatively small group (less than 8 individuals). The individuals had a wide diversity of backgrounds reflecting the interdisciplinary nature of remote sensing technology. Their formal education in remote sensing ranged from none to minimal with most of the people learning through on-the-job training.

Perhaps the strongest feeling expressed during the LACIE AI and DPA discussion was that there would be considerable mutual benefit derived from each group of analysts having a better appreciation of the other group's analysis techniques and job function. Although some specific suggestions were discussed, they did not differ substantially from those made during LACIE reviews and it was assumed these suggestions would be evaluated through the normal RID procedure. It was pointed out that LARS would be responsive to specific training requests associated with the LACIE or other JSC projects.

Training Program Design Activities

At the end of the first quarter of the contract period, it was recommended that two activities in the area of training programs and materials development be pursued during the remaining portion of the contract year. These recommendations were based on results of work carried out under the task analysis described above, an assessment of currently available training materials and identification of topic areas in which the development of new materials was needed.

The first recommendation was that a one to two week training session be held for 8 to 10 JSC personnel. The training session was to be held at LARS using existing slide-tape-study guide minicourses and videotaped instructional materials. Anticipating the availability of these materials to the remote sensing community, the training session was to emphasize: 1) the participants learning the subject matter and 2) preparing them to serve as instructors. It was felt that having training materials and a cadre of qualified instructor-consultants on site would improve and simplify on-the-job training of JSC personnel.

The second activity proposed in the first quarterly progress report was the development of additional training materials. Simulation exercises centered around the forestry applications project (FAP) and regional applications project (RAP) and additional titles in the FOCUS series were specifically recommended. Simulation exercises are designed to lead the reader through the professional thought and decision making processes typical of those required by a remote sensing analyst. The FOCUS series is a set of pamphlets which contains one page of text and one page of supporting figures.

Priority Assignment Activities

The following priorities were recommended in the first quarterly progress report for the training program and materials development work to be carried out during the remainder of the contract period:

1. Development of a forestry applications (FAP) simulation exercise.

2. Training 8-10 JSC personnel at LARS as described in the previous section.

3. Produce additional titles in the FOCUS Series.

4. Development of a simulation exercise for the regional applications project (RAP).

Work was pursued according to these priorities for the rest of the contract period. Results obtained in these areas are reported in the following sections.

Materials Development

The forestry applications simulation exercise was completed during the third quarter of the contract period and was submitted as an interim technical report (LARS Information Note 012376). This report documents in detail the analysis of a LANDSAT data set using computer-aided analysis techniques. It is designed to enable the reader to gain an appreciation of the decisions and trade-offs made by the experienced analyst. The publication describes the sequential process of analyzing a LANDSAT data set and emphasizes the interaction between man (analyst) and machine (computer). Typical products (results) of each step in the analysis are shown in the report.

The analysis described in the forestry application simulation exercise is shown in Figure 2.5-1. A better appreciation of the approach taken in the simulation may be obtained from reading the following paragraphs which have been adapted from the Overview section of the simulation:

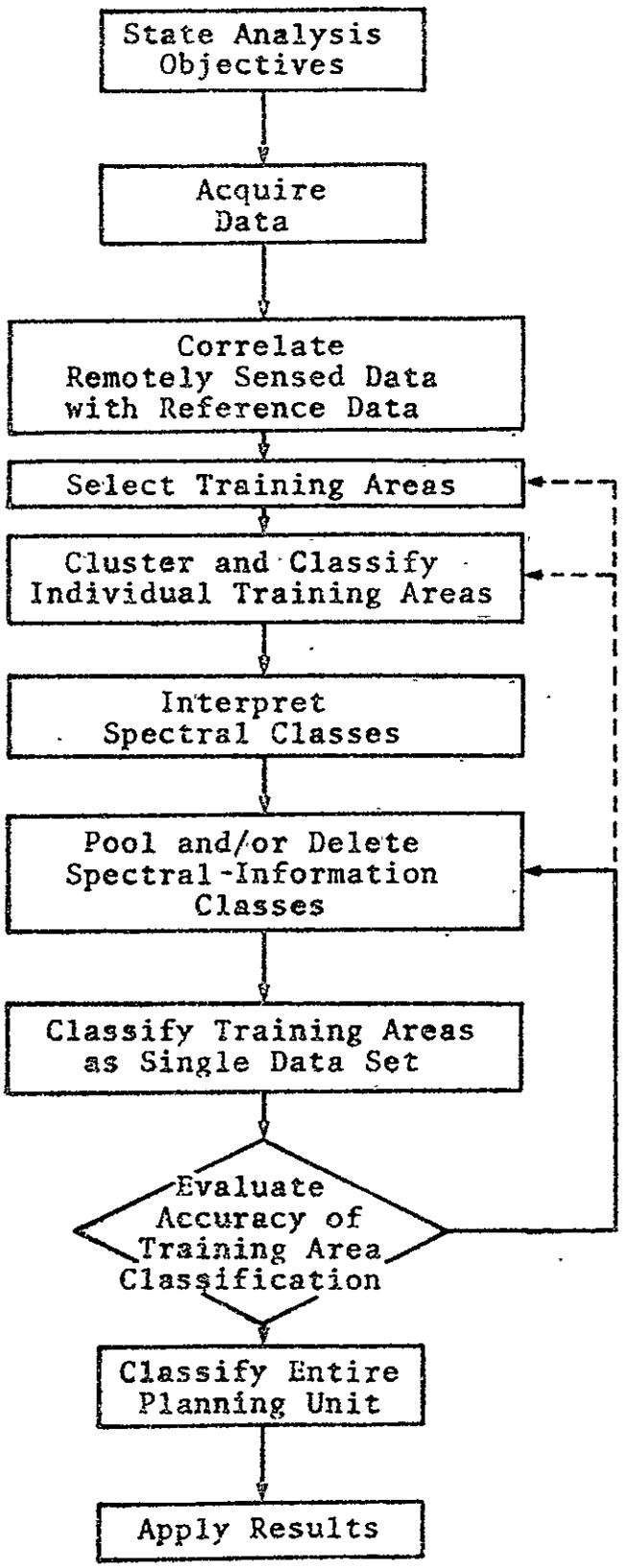


Figure 2.5-1 Typical Numerical Analysis Flowchart for Forestry Applications

The numerical analysis of remotely sensed data is a dynamic process which requires an interaction between man (analyst) and machine (computer). The process is both an art and a science, relying upon judgements and insights by the analyst as well as a documented technology of remote sensing analysis. A typical analysis sequence is shown in Figure 2.5-1. Even though it is shown here as basically a linear process, all of the steps are interconnected. At any step in the analysis, interpretation of the results of that step can lead the analyst to conclude that he should go back to a previous step and revise his procedure. For simplicity only the most commonly followed analysis sequence is shown.

Remote sensing techniques allow you to "survey" large areas with a minimum amount of time and cost. The computer can be "trained" to produce general land use maps as well as general forest cover maps. Even finer breakdowns of cover types may be achieved, such as timber stand maps, although mapping reliability is lower for these relative to general land use maps.

The first step is to state the analysis objectives. To do this, you must determine the geographic area of interest, the general cover types and the nature of the application to which the results will be applied. An additional component which is often included in the analysis objective is the expected classification accuracy for initial estimates of timber resources. An example would be to "determine the percentage of Hoosier National Forest in each of these cover types: conifers, hardwoods and other with 85% accuracy".

Next, the remotely-sensed data are correlated with the available reference data. The multispectral-scanner data may be from aircraft or satellites, such as LANDSAT. The reference data might include USGS topographic maps (quad maps), stand compartment type maps and related information, aerial photographs, U.S. Forest Service land use maps and actual ground observations. Each LANDSAT satellite covers the entire earth every eighteen days, so the analyst can most generally choose the time of the year most suitable for mapping the cover types of interest. The analysis sequence described in this simulation uses LANDSAT data.

The training areas are then selected. The training areas contain typical examples of each cover type of interest and are supplied to the computer in order to "train" it to classify unknown data points. There are some general selection criteria to aid the analyst in choosing training areas, but successful training area selection relies heavily on the analyst's previous experience and knowledge of the areas being studied.

When training areas have been selected, the next step is to use a computer processor (algorithm) called CLUSTER on each of the training areas individually. The CLUSTER processor uses information from more than one channel, or wavelength band, to produce a single computer-generated image. Since more information is used, the boundaries of ground features and cover types are usually more distinct on images produced by the clustering process than on a single-channel image.

After clustering, obtaining statistics, and classifying each of the training areas, the analyst looks at the output to see what each spectral class of the training areas represents. The spectral classes are groups of data points with similar spectral values (brightness levels). Aerial photographs and other reference data aid the analyst in making these associations between spectral classes and various cover types.

On the basis of the spectral separabilities and the known cover type information, the spectral-information classes may be pooled (merged together) or deleted. The spectral classes that are informationally and numerically similar (i.e. spectrally inseparable) are combined, while the spectral classes that are a mixture of two cover types (such as pasture and forest) may be deleted. The analyst should go back to his analysis objective(s) to help him decide which classes to combine and which to delete.

To check how well he did in the pooling and deleting of spectral classes, the analyst then classifies all the training areas together as a single unit. He then looks at the classification maps and compares them with other reference data. This step along with the output of the computer allows him to predict the probable accuracy to expect when he classifies the total planning unit.

With the output from classifying the training areas as a single data set, the analyst must predict if the training areas selected are going to allow him to meet his objective(s) when he classifies the total area under consideration. Will the classification yield the stated accuracy? Are all cover types adequately represented? If not, he must go back to previous steps as shown by the arrows on the right side of the flowchart. If possible, he may merely go back to "pool and/or delete" again. In some cases, he might go back and reselect the training areas. He may even need to go back to the beginning and restate his analysis objectives.

When he is satisfied with the classification data from the combined training areas, the analyst instructs the computer to classify the total area. Using pattern recognition algorithms, the spectral responses of each data point are "compared" to the training sample for each class, and the point is assigned to the "most likely" or most similar class. The output after this step

can be maps and data tables showing acres (hectares) for the mapped cover types.

As indicated earlier, numerical analysis of multispectral scanner data is a dynamic process with each step providing feedback to the previous step. For simplicity, the process is shown here as a linear sequence. In reality, the analyst has all steps in mind before he actually begins an analysis. He may also refer back to previous steps and modify his procedure as the analysis continues.

Following this overview the simulation exercise discuss each step in the analysis in detail with specific attention being paid to describe the analyst's reasoning and the basis of his decisions.

Work on a second simulation exercise began during the latter part of the third quarter and has continued throughout the fourth quarter of the contract period. The data set used for this simulation exercise has been drawn from among those analyzed in conjunction with the Regional Applications Project task of this contract. A first draft of the simulation exercise has been completed and it has undergone internal review. Based on the review the document is being revised to place greater emphasis on the underlying analysis principles and incorporate examples of typical output products into the document. Current status is that the revision is approximately 60% complete. After revision the document will again be reviewed internally and undergo student tryout prior to publication.

Three titles have been added to the FOCUS series during the contract year. Like other titles in the series these two-page foldouts consist of a diagram or photograph and an extended caption of three to four hundred words treating a single topic in remote sensing. They have been found to be especially useful for general briefings or introductory treatments of remote sensing topics. The new FOCUS titles along with a brief summary are:

LANDSAT Multispectral Scanner Data -

LANDSAT data in four spectral bands is available on computer compatible tapes and in various image formats. An example of an annotated LANDSAT image produce is given.

Clustering -

In remote sensing, clustering is used to determine the "natural structure" of data. It can be used to decompose complex data sets into simpler subsets and to determine data classes based on spectral rather than informational variations.

How the Earth Reflects -

Energy reflected by materials on the earth's surface varies according to the structure of the materials themselves and their conditions. Spectral differentiation is possible because vegetation, soil and water reflect energy differently from each other and because sub-categories of these materials

demonstrate spectral variations as well.

In addition to these additions to the series three FOCUS issues prepared during 1973 were reviewed for currency. They were judged to be satisfactory and not in need of revision at this time.

Interim Training Programs

One of the recommendations resulting from the task analysis conducted at the beginning of the contract year was to conduct a training session for 8 to 10 JSC personnel. The objectives of the training session were 1) to give the participants the opportunity to learn and work with the fundamental principles of remote sensing, especially those associated with the applications of pattern recognition techniques to the analysis of multispectral scanner data, and 2) to prepare the participants to serve as instructor-consultants for formal training programs or on-the-job training of personnel at JSC. Although no request for such an interim training program was received to a large extent the first of these objectives was met through the mechanism of JSC based personnel attending the Remote Sensing Technology and Applications short course sponsored by Purdue's Division of Conferences in cooperation with LARS. Those attending the short course from JSC during the past year were:

Name	Affiliation	
Richard Moke	NASA	September, 1975
Donald Saile	LEC	November, 1975
Milton Bertrand	LEC	January, 1976
Donald Hay	NASA	January, 1976
Olav Smistad	NASA	January, 1976
F. C. Kuo	LEC	January, 1976

To better facilitate our ability to respond to requests for training programs which might arise in the future, education and training materials which have been developed at LARS under NASA and Purdue University support have been organized and summarized in LARS Information Note 052576, Matrix of Education and Training Materials in Remote Sensing. This document, which will be submitted as a separate volume of this final report, organizes the remote sensing education and training materials developed by LARS in a matrix format. Each row in the matrix represents a subject area in remote sensing and the columns represent different types of instructional materials. This format proves to be useful for showing in a concise manner the subject matter content, prerequisite requirements and "technical depth" of each instructional module in the matrix.

Following a general description of the matrix and the instructional format of the materials in the matrix there are three examples of training programs designed to meet specific training objectives. Content of the programs were selected to match both the educational and experiential backgrounds of the participants and the constraint imposed by the amount of time available for training. The training program examples are followed

by a "catalog" of the instructional modules which contains a summary paragraph, list of recommended prerequisites and any special equipment or instructional aids which may be required to use the modules.

Availability of this document will facilitate implementation of remote sensing training requirements at JSC.

Remote Terminal Support

LARS continued its support of the remote terminal at JSC throughout fiscal year 76. Only a few problems with the equipment were reported (primarily the IBM 2780 Card reader/printer/punch and the Codex modems) but these were handled via calls to the appropriate vendor.

The JSC terminal maintained nine active computer ID's for most of the year, with usage displayed in Table 2.2-1. Note that very little use was made of the batch system this year.

The terminals were used for training new analysts, for general research and technology development, and very heavily for LACIE support. This latter support primarily made use of CMS capabilities and not LARSYS. People on the LACIE project used the terminals to statistically check the quality of training fields. This use of the terminals is evident in Table 2.5-2 where the total CPU time for LARSYS is much lower than the total CPU time used. Table 2.5-2 shows totals for CPU time attach time and CPU time used for LARSYS functions by JSC remote terminal users.

Another significant remote terminal activity was publication of the Remote Terminal System Evaluation reports (LARS Information Note 062775). Last year's final report stated that a draft version was being reviewed by NASA and early in the contract year the report was released for publication. The report documents the development of an earth resources data processing system which is being used by both LARS personnel and remote terminal users. Its value us a system for training, technology transfer, and data processing are evaluated in the report.

SUMMARY

Work in the area of technology transfer and remote terminal support was carried out in accordance with the task analysis, recommendations and priorities established during the first quarter of the contract year. In the area of training materials one simulation exercise was developed, significant progress has been made on a second analysis simulation, and three additional FOCUS issues were written. While no training programs were requested under the SR&T contract, 6 JSC based personnel attended the Purdue sponsored remote sensing short course during the past year. A report was prepared which summarizes and catalogs the remote sensing education and training materials developed at and available from LARS. This document will facilitate the planning and implementation of future training programs. The JSC-LARS remote computer terminal continued to be supported throughout the contract year.

Table 2.5-2 JSC Remote Terminal

CPU Time Usage (June, 1975 - May, 1976)
(Hours)

Computer ID		1975 June	July	Aug.	Sept	Oct.	1976 Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
JSC100	I	2.0	2.3	2.0	.9	.1							
	B	.0	.0	.0	.0	.0							
	T	2.0	2.3	2.0	.9	.1							
JSC102	I	.8	3.1	1.6	4.3	.35	.1	1.3	.6	.2	.1	.1	.1
	B	-	-	.0	.0	.0	-	.0	.0	-	-	-	-
	T	.8	3.1	1.6	4.3	.35	.1	1.3	.6	.2	.1	.1	.1
JSC308	I	1.4	3.9	3.4	0.5	.8	.3	.3	1.2	.5	1.2	.4	.5
	B	.0	.0	-	-	-	-	-	.0	-	.1	.0	-
	T	1.4	3.9	3.4	0.5	.8	.3	.3	1.2	.5	1.3	.4	.5
JSC404	I	.0	.0	2.5	1.1	.3	1.0	.3	.2	.0	-	-	-
	B	-	-	-	-	-	-	-	-	-	-	-	-
	T	.0	.0	2.5	1.1	.3	1.0	.3	.2	.0	-	-	-
JSC444	I	-	-	-	.2	1.2	.1	.2	.5	.2	.4	.5	.9
	B	-	-	-	-	-	-	-	-	-	-	-	-
	T	-	-	-	.2	1.2	.1	.2	.5	.2	.4	.5	.9
JSC500	I	-	.0	.0	.0	.0	.1	.0	.0	.3	.0	-	-
	B	-	-	-	-	-	.0	-	-	-	-	-	-
	T	-	.0	.0	.0	.0	.1	.0	.0	.3	.0	-	-
JSC600	I	1.8	1.7	2.6	2.1	.5	.7	1.7	1.9	1.5	2.0	.9	1.8
	B	-	-	-	-	-	-	-	-	-	-	.0	-
	T	1.8	1.7	2.6	2.1	.5	.7	1.7	1.9	1.5	2.0	.9	1.8
JSC601	I	-	-	-	-	.3	.1	.0	.2	.2	.1	.6	.6
	B	-	-	-	-	-	-	-	-	-	-	-	-
	T	-	-	-	-	.3	.1	.0	.2	.2	.1	.6	.6
JSC602	I	-	-	-	-	.5	1.0	.8	.7	.4	.2	.8	1.5
	B	-	-	-	-	-	.0	.0	.0	-	.0	.0	-
	T	-	-	-	-	.5	1.0	.8	.7	.4	.2	.8	1.5
JSC800	I	-	-	-	-	.7	4.6	6.6	.9	1.8	6.3	2.5	1.1
	B	-	-	-	-	.0	-	.0	-	-	.0	.0	-
	T	-	-	-	-	.7	4.6	6.6	.9	1.8	6.3	2.5	1.1

I = Interactive Terminal Use
B = Batch
T = Total

Table 2.5-3 JSC Remote Terminal

Monthly Computer ID Totals

	1975							1976				
CPU time	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Interactive	6.00	11.00	12.20	9.19	4.72	7.76	11.26	6.10	5.07	10.25	5.72	7.91
Batch	.01	.02	.01	.01	.01	.00	.00	.00	.00	.13	.03	.01
Total	6.01	11.02	12.21	9.20	4.72	7.76	11.26	6.10	5.07	10.38	5.74	7.91
LARSYS CPU time (hours)	1.05	2.81	2.86	4.42	.65	.02	1.38	1.19	1.41	1.51	.84	.01
Attach time (hours)	311	382	354	300	192	218	281	274	289	323	280	264

RECOMMENDATIONS

It is recommended that the regional application project simulation exercise already started be completed and that it and the forestry application simulation be evaluated and tried out by students at both JSC and LARS as part of the FY 1977 SR&T technology transfer task. The effectiveness of this instructional format should be evaluated and, if appropriate, suggestions for additional simulation exercise topics be made.

It is recommended that JSC based personnel with education and training responsibilities be throughly familiarized with the material described in Matrix of Education and Training Materials in Remote Sensing (LARS Information Note 052576) and a mechanism be established for evaluating the use of this document for the design of individual training programs.

A critical examination of the instructional materials matrix in LARS Information Note 052576 should be made by JSC and recommendations given to LARS concerning those areas in which additional materials are needed to meet the training needs of JSC.

The JSC-LARS remote terminal arrangement should be studied to see if any improvements in the terminal facility or its use should be made.

ACKNOWLEDGEMENTS

The project director for this task was Dr. John C. Lindenlaub. Major contributions to the project were made by: Shirley M. Davis, Dr. James D. Russell, Paula A. Pickett, Susan K. Schwingendorf and Bruce M. Lube.

2.6 Research in Remote Sensing Technology

INTRODUCTION

The Research in Remote Sensing Technology task consisted of three major activities with an additional image registration task included. The primary research area identified in the Statement of Work was termed "Ancillary Data Registration." The interest in this problem arises, for example, from the need of crop classification researchers to relate the spectral variations observed to causitive factors in the physical environment. Factors such as soil properties, land use patterns, precipitation, slope, thermal history, etc., are desired in digital form in registration with the multispectral data. This task was established to develop the techniques necessary to provide the desired ancillary data channels.

The second task consisted of further evaluation of an image enhancement algorithm developed in CY75 and reprogramming of the algorithm for improved efficiency. The algorithm both interpolates and enhances using an optimal instantaneous field of view filter[1] and shows promise for aiding visual interpretation of LACIE segment images. Part of the evaluation of the enhancement algorithm included classification accuracy experiments and these results are not yet available.

The third research task consists of a spectral reflectance study of soils using the EXOTECH field spectrometer. The task seeks to relate the detailed soil reflectance spectrum obtained from this instrument to physical properties such as Cation Exchange Capacity, Organic Matter Content, etc.

An additional task which consists of temporally registering four passes of LANDSAT data for two LACIE intensive study sites is reported herein. This work was done with computer resources available under the 2.6 task.

Although the ancillary data overlay and image enhancement work was not directed to be in support of LACIE the results have turned out to be of potential benefit to the project. Signature extension research will be aided by the availability of registered physical data and segment visual interpretation may benefit from the enhancement algorithm. The temporal registration task is part of research on the effects of misregistration and is in direct support of LACIE.

Ancillary Data Registration

An important motivation for ancillary data registration derives from the signature extension problem. Physical

factors are assumed to cause variation in spectral response which invalidates training parameters at observation points distant from the training area. Digital registration of physical factor data with remote sensing data will enable quantitative evaluation of the effect of these factors. The same reasoning applies to many other classification problems such as forestry where slope and elevation is important and for these reasons an ancillary data registration task was pursued.

The types of physical data which are most often encountered are in the form of maps. The data is usually originally obtained in tabular of photographic form and compiled into a map. The formats of the data on the map includes contours, lines, polygons and points. Figure 2.6-1 graphically describes some of these forms. It is assumed that the physical procedures for transforming the map into a digital image-like record registered pixelwise with remote sensing data. The basic steps in the map transformation process are assumed to be:

1. Digitization of the physical map.
2. Conversion of the lines and points to a uniform grid of points.
3. Registration of the gridded data to a reference data set.

Each step involves a number of parameters and algorithms. Some of the considerations which must be made are:

1. Density of samples for contours and arcs.
2. Effects of position errors.
3. Scheme for defining interiors of polygons.
4. General considerations of the information content (bandwidth) of a map.
5. Interpolation algorithms for gridding contour and point data.
6. Coordinate systems and warping functions for registration.

These factors were considered in the context of the variables of interest for spectral strata definition. The variables currently under study are:

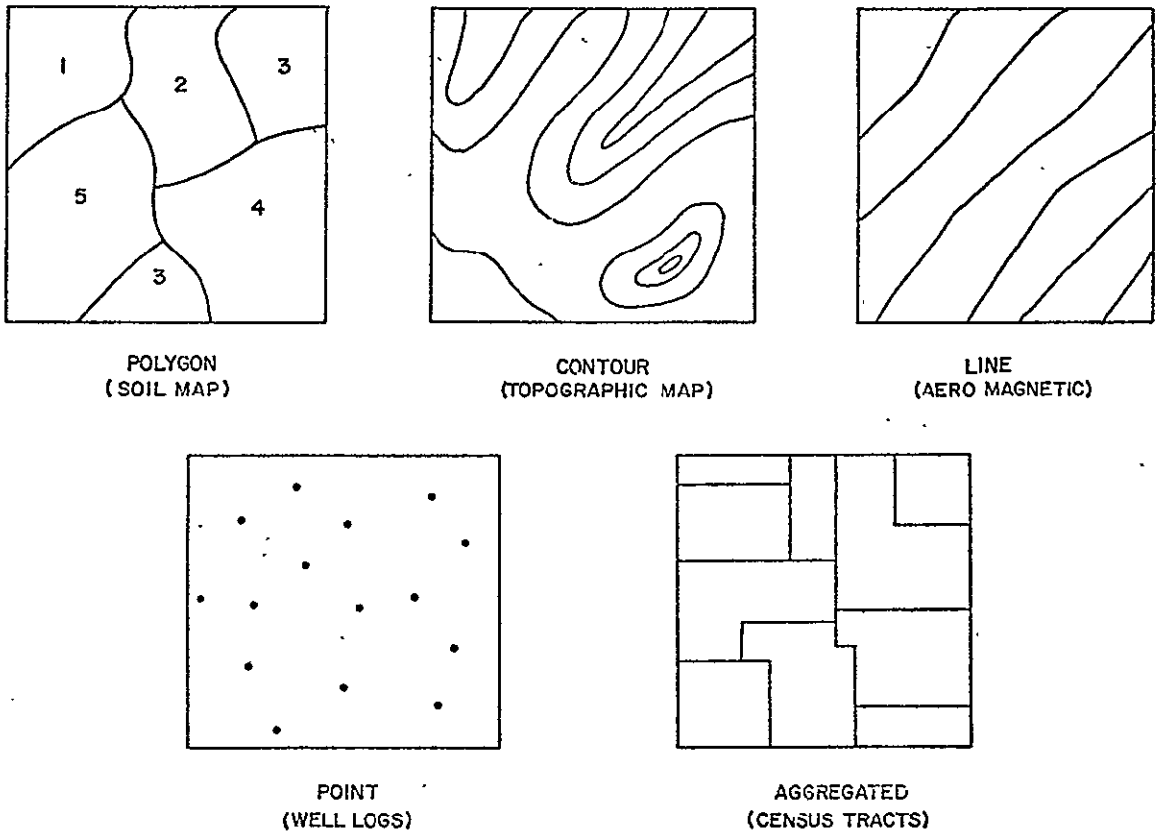


Figure 2.6-1. Ancillary Data Format Examples

1. Soil Type and Association.

2. Temperature and Precipitation.

3. Land Use.

The study first made some general considerations of the map data digitization problem. The polygon format requires the simplest data processing operations since constant values are all that need to be filled in inside the polygon and these values are specifically defined. The contour, line, and point formats require smooth surface interpolation between data points and these cases present the greatest challenge. The aggregated data case is also a relatively difficult problem in that the value given in a tract is a total of some variable such as population and dividing the total by the area and assigning the average to the entire tract may give misleading or undesirable weights to certain points in the tract. Curve fitting techniques may be desirable in this case which would be for example match densities at the boundaries of tracts. It was the original intention in the study to address each of these data types and research techniques for digitizing each form of data. The main application of the research was to be the signature extension or spectral strata task and the project ultimately focused on the requirements for this task. The data types considered for the spectral strata were all of the polygon type and the resources required to complete the work on the polygon were such that other types of data could not be considered during the year.

Thus, the ancillary data registration project considered four polygonal data sets (soils, land use, temperature, precipitation), developed procedures, and achieved registration of these four variables with a full frame of LANDSAT data in the project period. Data did not become available until the middle of the second quarter thus the period the overlays were completed in was slightly over two quarters. Each polygon data type presented different problems and each is discussed below.

Soil Map

The test sites of interest were in Kansas and included LACIE intensive study sites. It was decided that since the map data was available for the entire state the total area would be overlayed on LANDSAT frames as desired. The soil map used is described in the Spectral Strata Section (2.2) and a reproduction is presented in Figure 2.2-4.

The map was mounted on a coordinate digitizing table and each line separating different soil types was digitized by punching the coordinates of points along the lines on

LABORATORY FOR APPLICATION OF REMOTE SENSING
PURDUE UNIVERSITY
BOUNDARY RE-PLOT MAP: SOIL ASSOCIATIONS OF KANSAS

44 SHORT ARCS

ARC	LEFT	RIGHT	ARC	LEFT	RIGHT
5	0	1	10	0	2
14	10	5	15	4	5
16	10	5	19	11	4
25	0	1	27	0	1
29	0	4	33	11	0
34	11	12	35	10	12
37	7	12	40	00	00
41	07	08	42	7	1
44	1	0	52	11	14
55	10	15	58	27	15
61	14	10	75	24	25
75	25	29	90	30	20
99	35	21	100	34	21
105	20	25	111	19	25
117	10	20	130	40	42
132	37	0	135	40	0
139	51	2	150	32	0
139	31	2	141	31	30
149	32	30	160	33	30
156	31	30	156	31	37
163	37	34	105	40	30
105	41	30	104	0	2

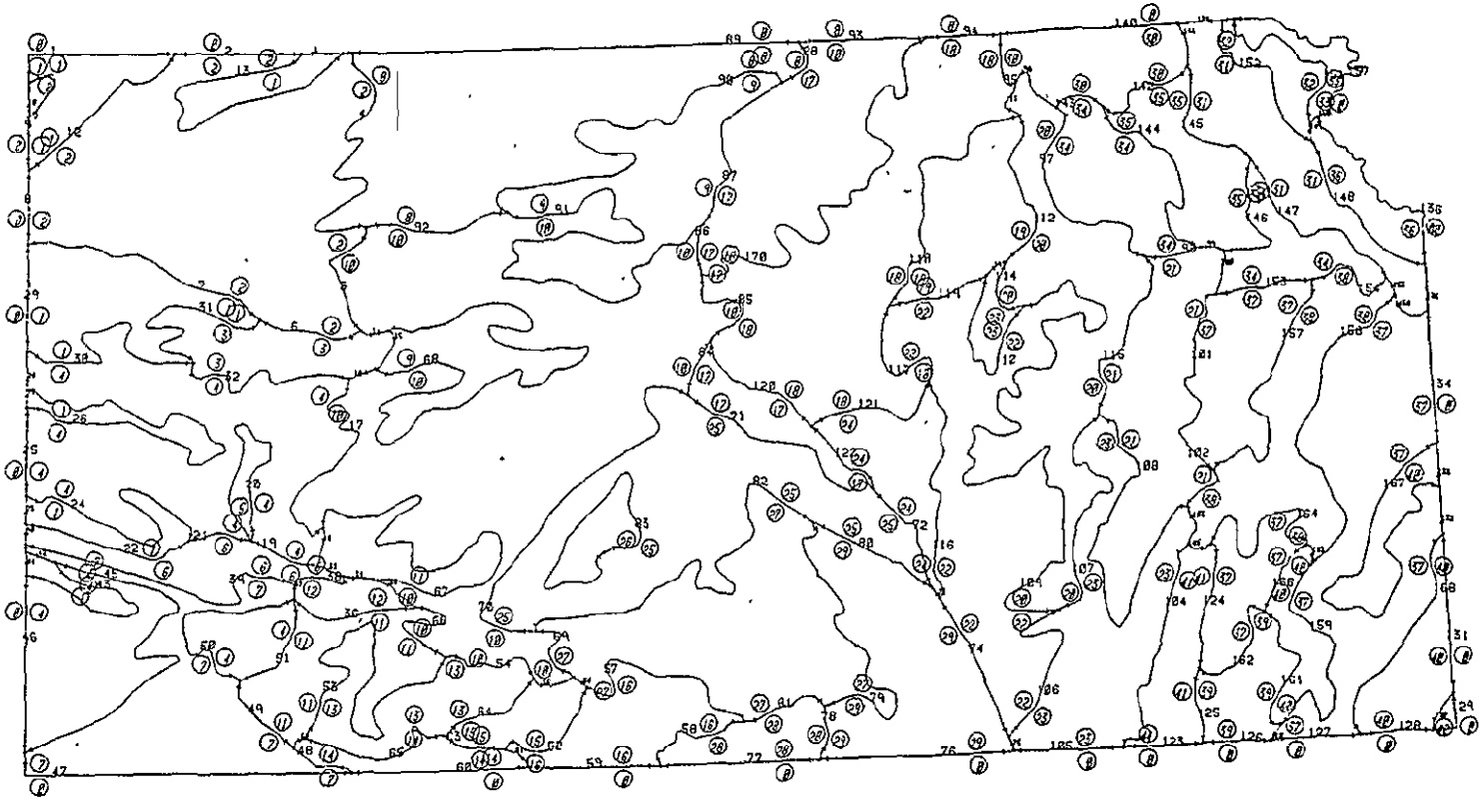


Figure 2.6-2 CALCOMP Boundary Plot for Checking Accuracy of Kansas Soil Map Digitization.

cards. The accuracy of the digitized points is nominally .01 inch and the point density used varies according to the curvature. Points were more widely spaced (nominally .2 inches) on low curvature areas and as little as .05 inch on sharply curving segments.

A line between two intersections with other lines dividing other soil types is called an arc. Each arc is digitized individually and assigned a number. The output from the digitizing phase is thus a set of cards with x-y coordinates for each arc and an identifying number for that arc. For the soil map 170 arcs were digitized.

The next problem was to identify the contents of the areas enclosed by the arcs. A procedure was established by which code numbers were assigned describing the contents of the area to the left and right of the arc where the forward direction is the direction the arc was digitized. The complete list of arcs with the left and right codes is then input to a boundary processing software package which produces a gridded image like data set for the soil variables. In order to check the accuracy of the arcs and their labels a CALCOMP plotting routine was assembled which draws the arcs and labels at the same scale as the original map. The plot can then be overlayed on the map and each arc visually checked. Figure 2.6-2 contains the arc plot for the soil map.

Concurrently control points were identified on the map and in the LANDSAT imagery to enable registration of the map data to the LANDSAT data. Road intersections, corners of water bodies, river bends, etc., were chosen and recorded. The locations of these points were digitized and punched on cards thereby defining the point in the digitized soil data. The points were also located in the LANDSAT data and the line column number recorded. The steps followed in the procedure developed digitizing this data are diagrammed in Figure 2.6-3.

Once the data and control points were all in digital form the gridded data set was generated. Arcs are sorted and code numbers assigned to each pixel according to which polygon the pixel is in. The result of this step is a digital image array having nominally the same scale as the LANDSAT data.

The final step in the process is registration to the LANDSAT data. The control points are used to estimate the coefficients of a bi-quadratic function which described the geometric transformation of the gridded map data to cause it to match the LANDSAT data. This function is then applied to the data producing a final output file as indicated in the lower right of Figure 2.6-3. The final output data set

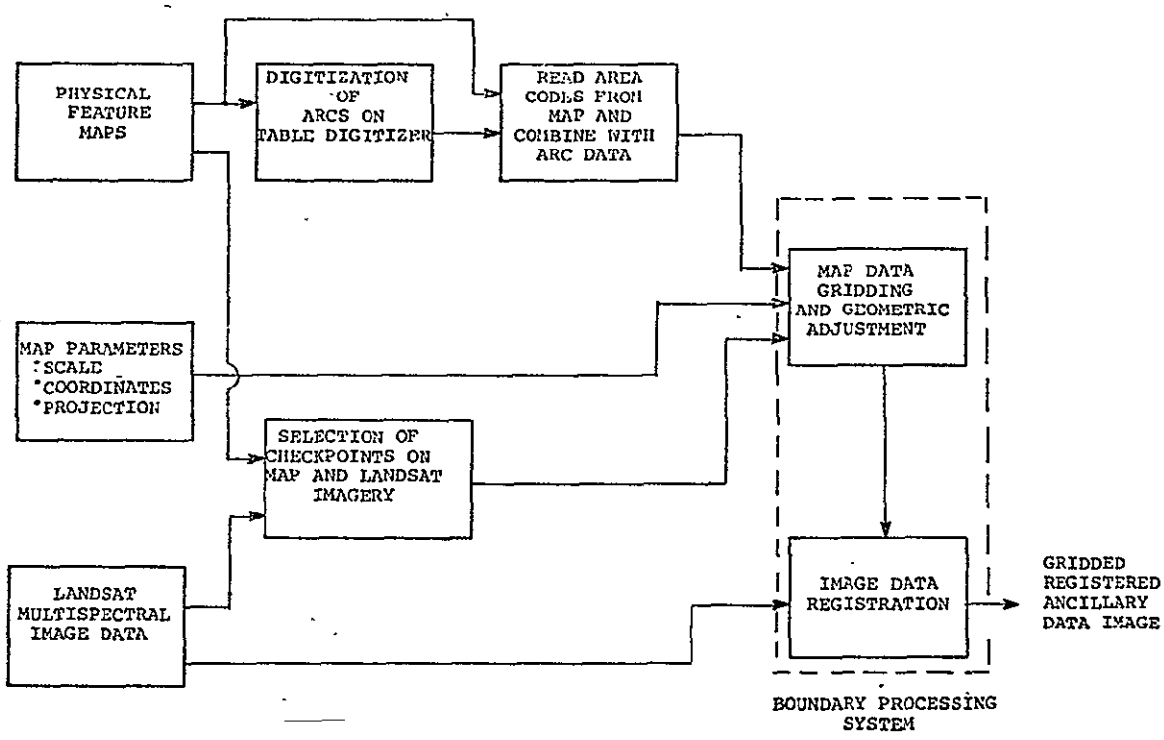


Figure 2.6-3. Functional Flow Diagram of Ancillary Map Data Digitization Procedure.

illustrated in Figure 2.2-5 in a previous section via a gray scale image.

The accuracy of the overlay depends on two aspects. The first is the accuracy of the physical maps themselves. The second is the accuracy of the registration process.

The accuracy of these processes thus depend on the following:

1) Uncertainty of digitization. The table digitizer has a resolution of 0.01 in. That means that there is an uncertainty of 0.01 times the scale of the topological map. For example, if the county map has a scale of 1,000,000, then the uncertainty is 0.01x1,000,000/12 ≐ 833 ft. If the pixel is 250 ft. wide, then digitization could cause an uncertainty of 3.3 pixels. It should be noted that we may reduce such uncertainty by using a more detailed map.

2) Uncertainty due to quantization and instability of table digitizer. The digitizer has three quantization levels: 100, 200 and 400 parts/in. However, since the resolution constraint by the moving arms and other mechanical parts is 0.10 in., using the high quantization scale does not improve accuracy but brings in the problem of stability. The digitizer drifts from time to time, i.e., such that the same values at the same point are not obtained after moving across the map, and be experience the drift is estimated to be about 2 to 4 units at the 200 pts/in scale. So the uncertainty contributed by instability is about 0.01 in., the same amount as that by digitization.

3) Uncertainty of location of checkpoints. Checkpoints are identified from the physical maps as well as from the multispectral image. Usually intersection of highways and other landmarks are used as checkpoints because these can be easily identified on the multispectral image. Both digitization of checkpoints on the map and visual location of checkpoints in the multispectral image cause uncertainty. Digitization uncertainty, as estimated above, is about 3.3 pixels; identification of checkpoints is assumed to within 1-2 pixels. Thus, total uncertainty for this part is about 5 pixels.

4) Uncertainty due to image registration. This by far has the least uncertainty. According to the way the boundary program is structured, the registration error is held to be within half a pixel.

The table digitizer is operated manually, however, and human error is likely to be more than .01 inch. It is hardly

possible for a human eye to resolve a distance of 0.01 in., in fact, it takes extreme concentration to distinguish a distance of 0.03 in. Thus, this is more likely the most serious factor degrading the accuracy of the overlay at these scales.

Land Use Map

A general land use map of the state of Kansas was the second physical data map processed. This map is described in Section 2.2 and illustrated in Figure 2.2-6 in that section. The same procedure was followed for this map as for the soil map and a numerical image data set was generated with land use codes from one to twelve. A considerable amount of difficulty was encountered due to the fact that there were over 1000 arcs in the land use map and the software had been set up for less than 1000. The CALCOMP replot of the land use map arcs is presented in Figure 2.6-4. This data was then registered on the LANDSAT frame in the same manner as before.

Temperature and Precipitation Maps

The temperature and precipitation data was available only on a county basis from World Meteorological Organization tapes as described in Section 2.2. Thus a constant numerical value was generated to fill in the area inside each county. The county boundaries were digitized and processed similarly to the previous maps and two different sets of numbers input for each county for temperature and precipitation. A boundary plot of the digitized county boundary is presented in Figure 2.6-5.

There are many ways the temperature and precipitation numbers could be defined. For the initial phase of the study the total winter wheat growing season average was used comprising the interval October 1973 through June 1974. Other combination of periods could be used in a number of additional channels for more detailed correlation studies.

The product of the first years effort in ancillary data registration is a set of procedures and a four channel overlay data set for one frame of LANDSAT data in Kansas.

The experience gained by taking one approach to the ancillary data registration problem and completing the processing of four variables was very valuable. This work can lead the way toward further research on this problem and hopefully more efficient methods will be developed as a result.

LABORATORY FOR APPLICATION OF REMOTE SENSING
PURDUE UNIVERSITY
BOUNDARY RE-PLOT MAP: LAND-USE PATTERNS OF KANSAS

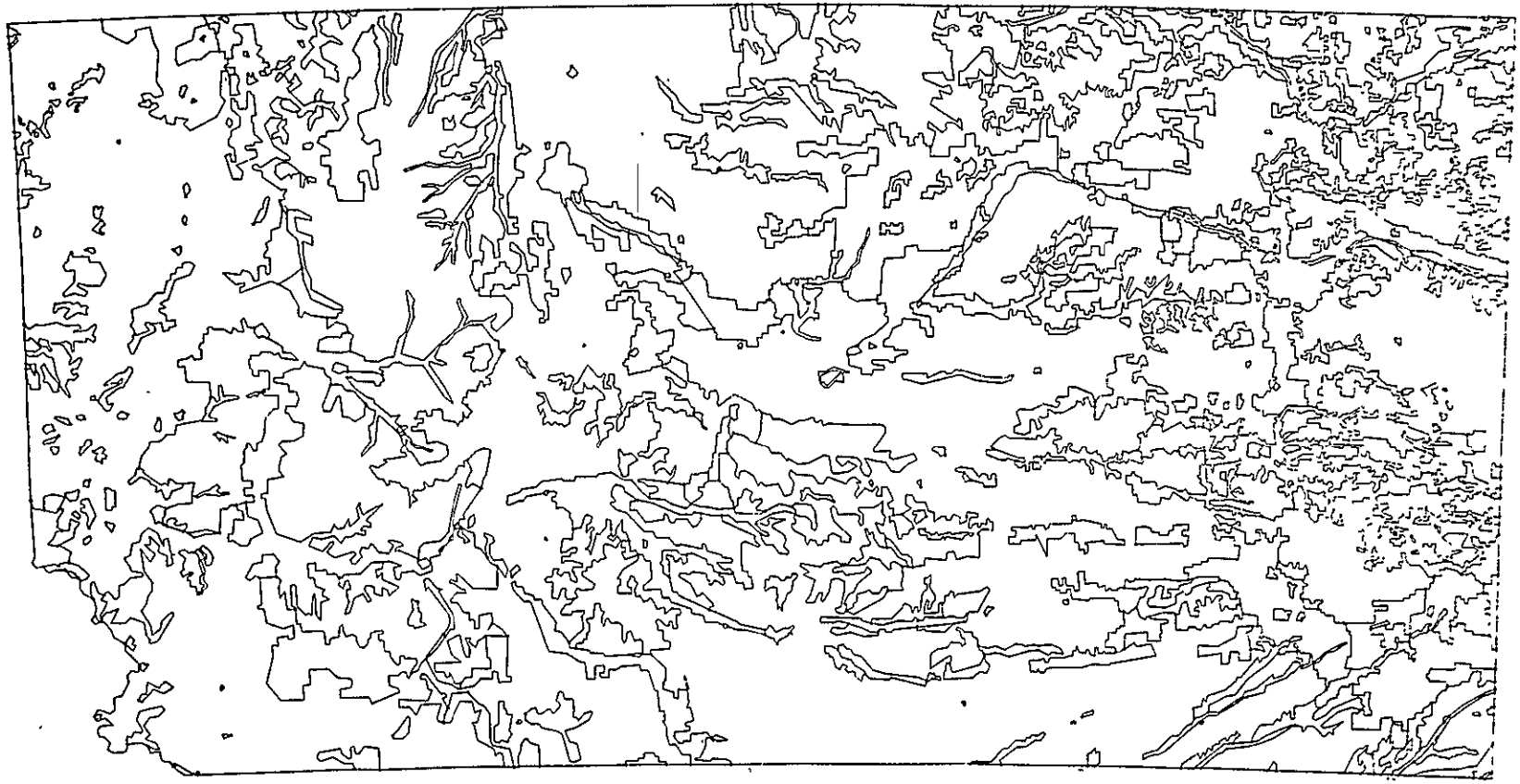


Figure 2.6-4 CALCOMP Boundary Plot for Checking Accuracy of Kansas Land Use Map Digitization.

LABORATORY FOR APPLICATION OF REMOTE SENSING
PURDUE UNIVERSITY
BOUNDARY RE-PLOT MAP · COUNTIES OF KANSAS

43 SHORT ARCS

ARC	LEFT	RIGHT	ARC	LEFT	RIGHT
5	2	8	6	3	7
12	4	6	18	62	6
19	61	12	21	11	5
25	11	7	29	18	8
33	7	16	36	11	15
40	11	15	45	58	18
46	41	15	48	28	15
53	19	15	56	18	16
66	21	25	76	58	22
79	58	24	97	31	59
107	55	57	111	53	59
115	48	58	118	43	58
133	43	47	141	51	49
147	53	46	148	54	52
167	48	58	187	68	62
171	59	57	174	58	78
178	58	66	181	64	58
184	62	64	187	61	63
217	76	82	252	61	64
272	72	60	289	60	87
302	28	24	305	28	26
305	80	60			

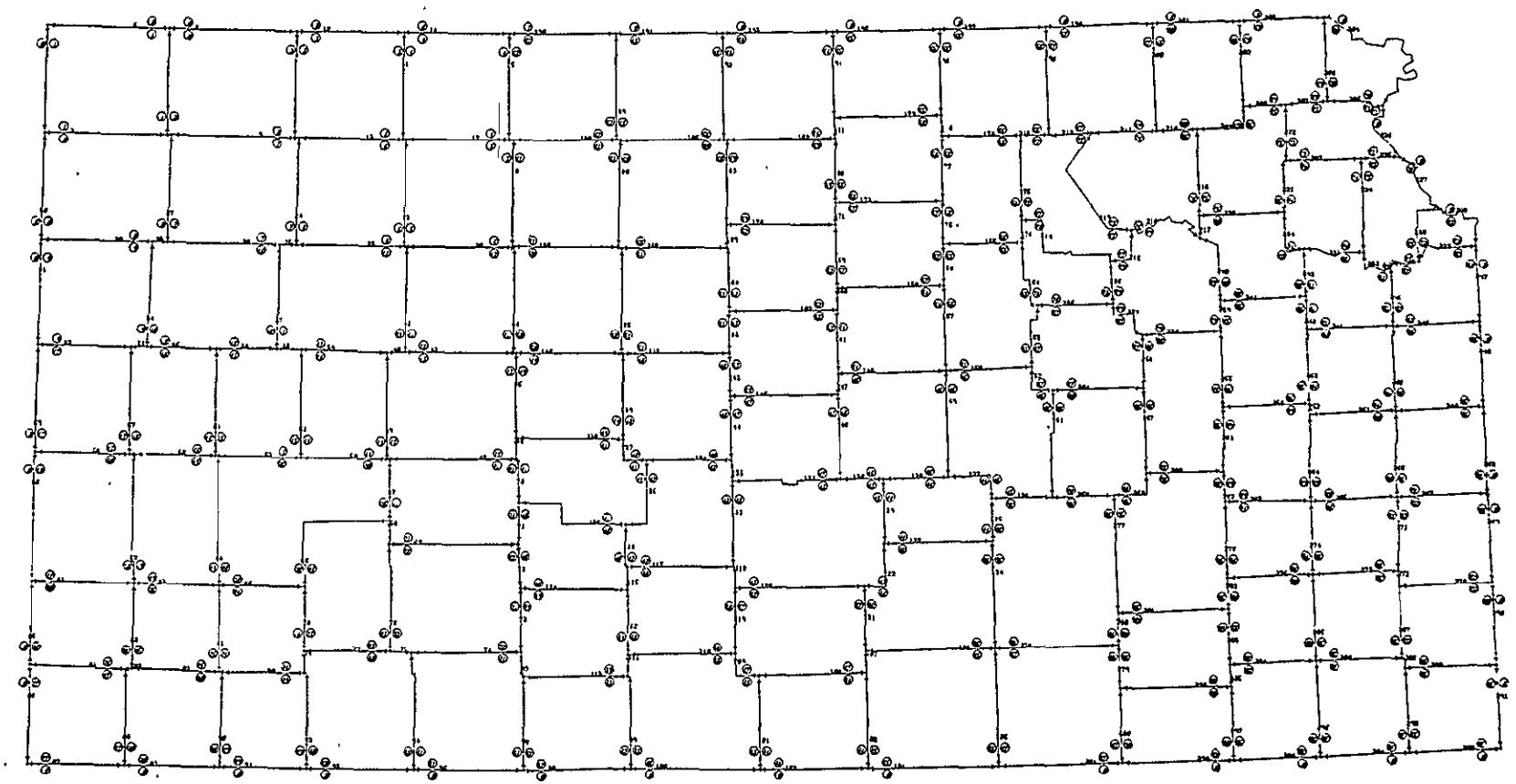


Figure 2.6-5 CALCOMP Boundary Plot for Checking Accuracy of Kansas County Boundary Digitization.

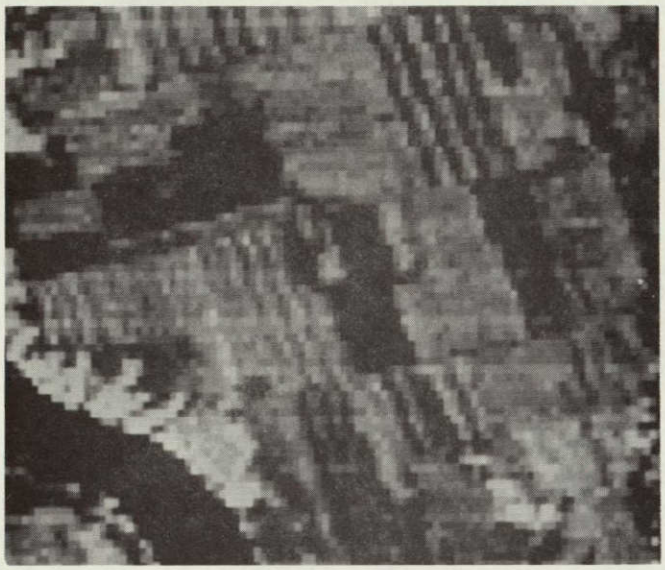
Image Enhancement Evaluation

The optimum IFOV (Instantaneous Field of View) enhancement filter developed in CY75 and discussed in previous reports was applied to several LACIE segments for further evaluation. In addition, a modified program for cubic interpolation was generated and applied to the same data. Also, an existing blocking expansion program was applied to the data sets. Thus, a comparison of blocking, cubic interpolation and IFOV filtering and expansion was obtained. The results using the IFOV indicate a significant sharpening while maintaining a "smooth" characteristic to the image as compared to blocking. The cubic causes considerable "blurring" and was considered not very desirable.

Four examples of the enhancement comparisons are presented in Figure 2.6-6, a-d. The top image is generated by duplicating original LANDSAT pixels by a factor of three horizontally and four vertically to compensate for the unequal sampling rates of the scanner. The typical blocking effect is seen here. The middle images are cubic interpolations obtained by fitting a bi-cubic polynomial to the sixteen points surrounding the new point being created and evaluating the polynomial at that point. Again a 3x by 4x expansion was created. The pixels at the original LANDSAT point locations have their original values. The bottom images are IFOV filter outputs obtained by interpolating and filtering the original data to produce a 3 by 4 expansion and enhancements. Here all original points have new values. The first site (Hill Co., Montana) covers smaller area (approximately 2 by 3 miles) than the other three which are LACIE segments. It was judged from the relatively low quality reproductions that the filtered data had the best visual quality of the three but reproduction on a high quality image writer should be carried out. LACIE analysts should make the comparisons to determine the usefullness of the enhancement in an operational environment.

To facilitate this comparison, six enhancement comparisons consisting of eighteen four band data sets were delivered to JSC in March 1976 for reproduction on the production film converter. Transparencies will be available for evaluation by both LACIE analysts and LARS personnel.

The IFOV enhancement program was completely reprogrammed during CY76 and the new version was used for these examples. The program performs a four channel enhancement nominally five times faster than the previous version. The previous version did only one channel and required extensive tape and disk buffering. The new convolution algorithm uses a multichannel cylindrical buffer and computes all points in one



Original in Color

Blocked Enlargement

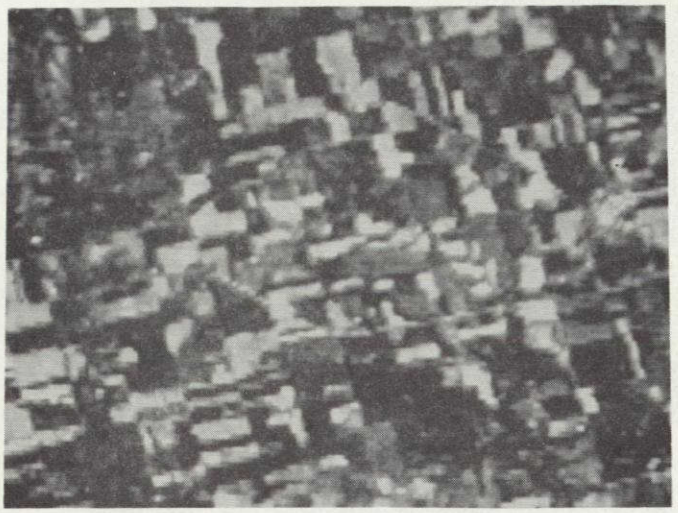
Original in Color

Cubic Interpolation

Original in Color

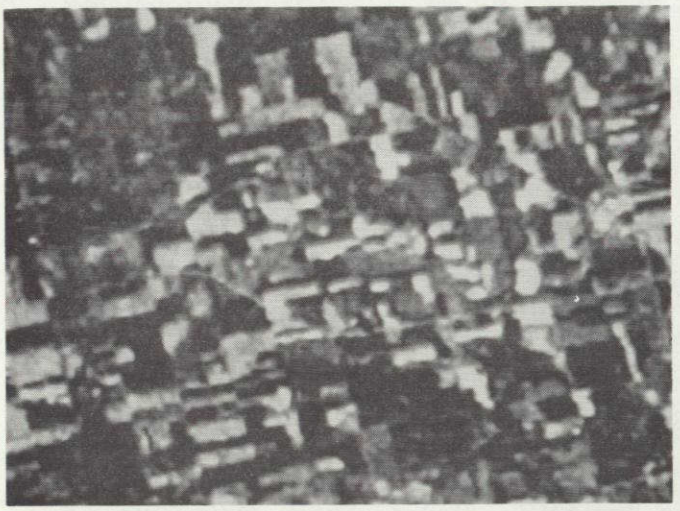
IFOV Filtering

Figure 2.6-6a Enhancement Comparisons for Run 73124700, Hill County, Montana, July 3, 1973.



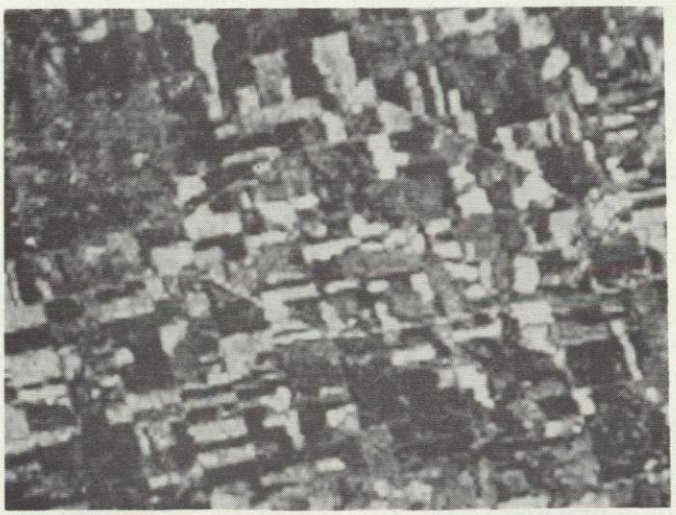
Original in Color

Blocked Enlargement



Original in Color

Cubic Interpolation



Original in Color

IFOV Filtering

Figure 2.6-6c Enhancement Comparisons for Run 74030600, McPherson County, Kansas, May 6, 1974.

Original in Color

Blocked Enlargement

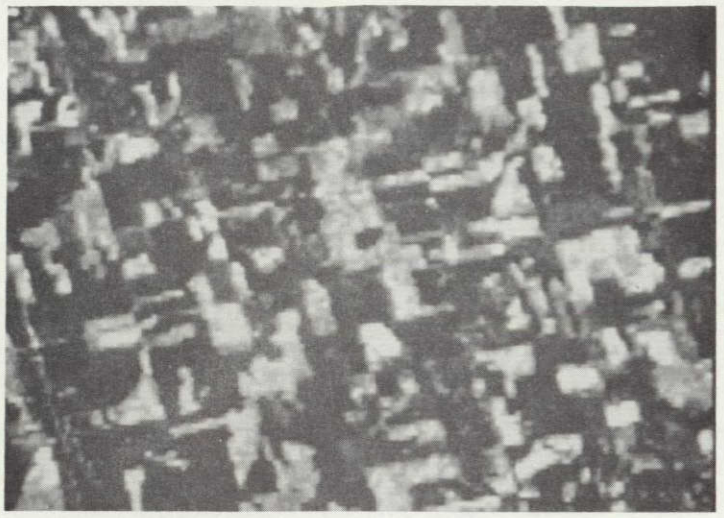
Original in Color

Cubic Interpolation

Original in Color

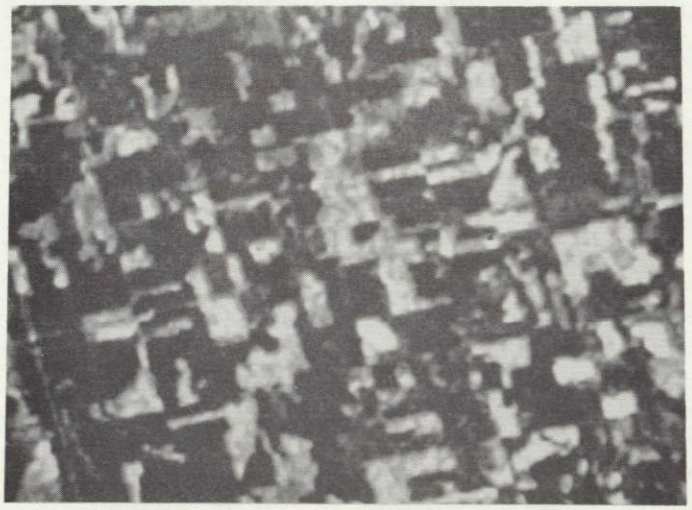
IFOV Filtering

Figure 2.6-6b Enhancement Comparisons for Run 74030400, Ellsworth County, Kansas, June 12, 1974.



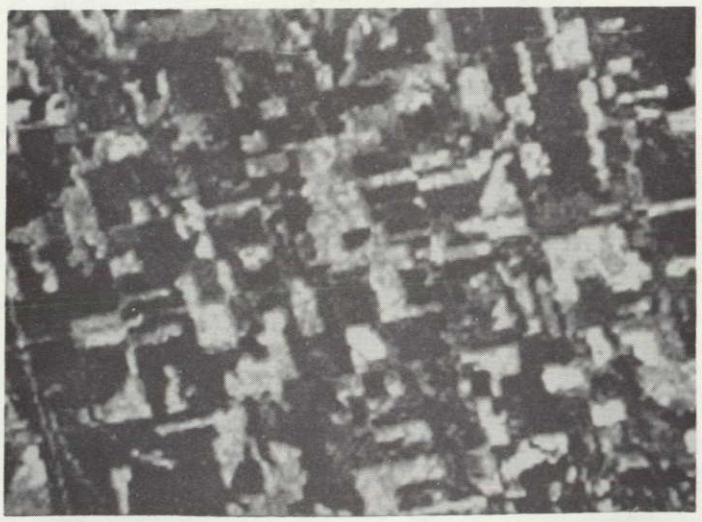
Original in Color

Blocked Enlargement



Original in Color

Cubic Interpolation



Original in Color

IFOV Filtering

Figure 2.6-6d Enhancement Comparisons for Run 74030200, Barton County, Kansas, June 12, 1974.

output data line in one pass rather than two.

The basic process of convolution as implemented in the current IFOV algorithm is to map each point of the input image to a four by three block of points in the output image. The points in the output are a weighted sum of the original point and four points to either side and above and below it.

Let $P_{0,0}$ be the original point; a $\hat{P}$ array will be generated as follows:

$$(P_{0,0}) \longrightarrow \begin{bmatrix} \hat{P}_{0,0} & \hat{P}_{0.1} & \hat{P}_{0,2} \\ \hat{P}_{1,0} & \hat{P}_{1,1} & \hat{P}_{1,2} \\ \hat{P}_{2,0} & \hat{P}_{2,1} & \hat{P}_{2,2} \\ \hat{P}_{3,0} & \hat{P}_{3,1} & \hat{P}_{3,2} \end{bmatrix}$$

$$P_{n,\ell} = \sum_{j=-4}^{4} [C^{(v)}_{n,j} \sum_{k=-4}^{4} C^{(h)}_{1,k} P_{j,k}].$$

The two matrices $C^{(v)}$ and $C^{(h)}$, represent the weights of input points in calculating the new point. $C^{(h)}_{1,k}$ represents the weight of $P_{j,k}$ in calculating an intermediate point $\tilde{P}_{j,1}$. $C^{(v)}_{1,k}$ represents the weight of $\tilde{P}_{k,j}$ in $\hat{P}_{1,j}$.

The two dimensional nature of convolution facilitates its partitioning into horizontal and vertical components. This horizontal process is repeated until sufficient $\tilde{P}$ lines have accumulated to allow the calculation of a set of $\hat{P}$ lines. The $\tilde{P}$ lines are kept in a cylindrical buffer. After each $\hat{P}$ calculation, the oldest line in the buffer is discarded and a new P line is read in. The buffer, therefore, moves like a rolling cylinder through the input image.

Each P can be calculated by the following:

$$\tilde{P}_{1,k} = \sum_{j=-4}^{4} C^{(v)}_{1,j} \tilde{P}_{j,k}$$

Each $\hat{P}$ can, in turn be calculated by:

$$\hat{P}_{1,k} = \sum_{j=-4}^{4} C^{(v)}_{1,j} \tilde{P}_{j,k}$$

In the flow chart in Figure 2.6.7 INLINE is the P line, CYLINDER represents the P lines and OUTLINE represents the $\hat{P}$ lines. NSAMPLES is the number of points per P line.

Directly computing $\hat{P}$ and P would require the following time (ignoring stores and loads):

$$T_D = 9 \cdot (t_a + t_m) \cdot 9 \cdot n_{lines} \cdot n_{columns} \cdot 12$$

$$= 972(t_a + t_m)\, n_{lines} \cdot n_{col}$$

where t_a is CPU time for addition,

t_m is CPU time for multiplication,

n_{lines} is number of input lines,

and $n_{columns}$ is number of input columns.

Partitioning the process into components gives the following time:

$$T_p = 3 \cdot n_{lines} \cdot n_{cols} \cdot 9(t_a + t_m) + 4 \cdot (3 \cdot n_{lines} \cdot n_{cols}) \cdot 9(t_a + t_m)$$

where the terms represent horizontal and vertical times respectively.

$$T_p = 135(t_a + t_m)\, n_{lines} \cdot n_{col}$$

Computing the percentage improvement gives the following conclusions:

$$I = 100\% - \frac{T_p}{T_D} \cdot 100\% = 100\% - 13.7\%$$

$$= 86.3\%$$

In summary, the convolution algorithm generates for each input point a 4 by 3 block of output points. Each output point is a weighted sum of some group of input points. By calculating an intermediate group of horizontally computed points a new savings of 86.3% over directly computing each output point can be recognized in computer time. The 86.3%

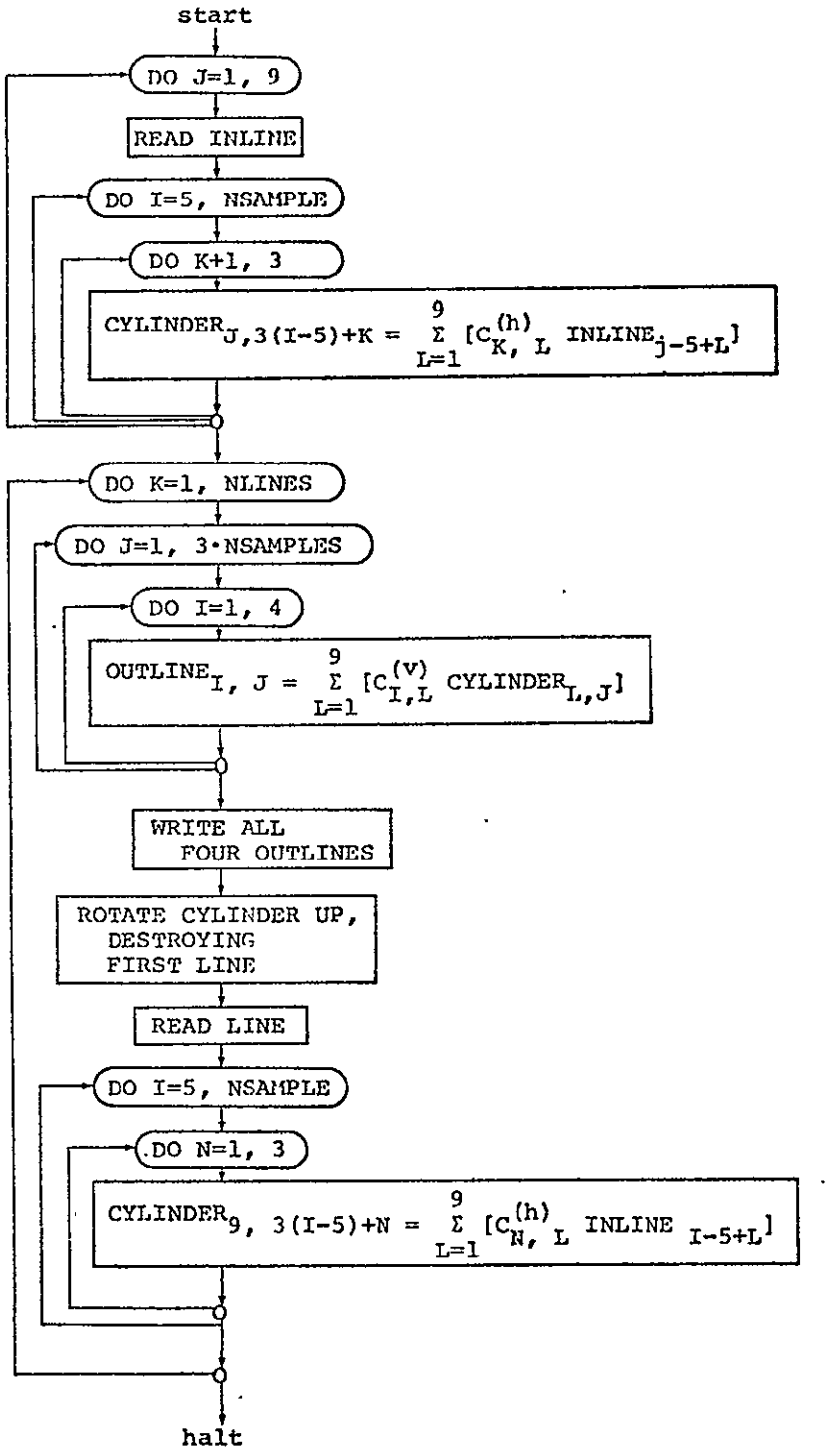


Figure 2.6-7 Flow diagram for IFOV enhancement algorithm.

improvement is in the actual mainframe or CPU time (e.g., if the original algorithm took 100 sec. CPU time the new one would take 13.7 sec.).

A cost comparison was made while computing the enhancements although the programs are not optimized to the extent they could be. Significant improvements are expected with the specialized programming approaches. For a 192x192 point area expanded 3x4 to a 576x768 point area in 4 channels by each of the three algorithms the CPU computing time on an IBM 360 Model 67 computer was 15 sec. for blocking, 338 sec. for cubic and 965 sec. for IFOV filtering. The high cost of IFOV is due to the fact that 81 original points are used to compute each of the new points where with cubic only 16 are used.

The complete IFOV enhancement program in source listing form is reproduced in Appendix C for purposes of documenting the final result of the IFOV filtering research program. Further comparison enhancements for blocking, cubic, and IFOV filtering can be run at LARS until such time that JSC chooses to execute the program on their system.

Temporal Registration of LACIE Sites

JSC requested that LACIE data for two sites and four dates be registered by the LARS system as an additional task. LARS agreed to do this and the registrations were carried out by the end of the CY. Both nearest neighbor and cubic interpolation resampling was used as requested.

Spectral Reflectance Properties of Soils

Remote sensing of both terrestrial and extraterrestrial environments has become a common scientific activity (Hunt and Salisbury, 1970). In this effort, spectroscopic techniques are often used to determine the composition and nature of the target material, whether the target is a pine forest or the surface of the moon. In most cases, the spectral behavior of rocks or the soils derived from them is of interest, either as the target spectral response itself, or as a background that must be understood in order to define its effects or to be eliminated.

Despite widespread use of spectroscopic techniques in remote sensing, the spectral behavior of soils is not well understood. Too often an empirical approach to a specific remote sensing problem has been used; detailed studies of spectral behavior have been done in a context inappropriate for remote sensing use; or the spectral data obtained in

a study have been too voluminous and difficult to organize for publication.

This research is a continuation of work devoted to a study of the relationships between the physico-chemical properties of soils and their spectral reflectance in the visible and infrared portions of the electromagnetic spectrum. The interpretations of these relationships are presented in such a way as to be useful to those engaged in spectroscopic remote sensing.

Using the BRF/Spectroradiometer (Exotech Model 20C), spectral reflectance values in the wavelength region 0.48 to 2.38μm were generated for the 56 samples of benchmark soils representing the different climatic regions of the United States (Koppens Classification). These soils also represented eight of the soil orders as defined by the Soil Taxonomy.

Data for 9 measured and 5 calculated properties of these samples were analyzed by stepwise multiple regression analysis. Twenty-one biochemical measurements were made on 27 of the 56 samples in order to study the relationships between organic constituents and spectral reflectance.

Spectral reflectance values at selected wavelength intervals were designated as the dependent variables and the physicochemical properties were the independent variables. The physicochemical properties of the soil having the highest correlation with spectral response are: cation exchange capacity, and the contents of silt, clay, iron and organic matter. Silt content was the single most significant parameter of those being studied in explaining the spectral variations of soils having r-values ranging from 0.41 to 0.62 and 0.63 to 0.70 in the visible and infrared regions, respectively. Organic matter content contributed significantly ($r = .30-.53$) to the explanation of the variation in spectral response in the visible region.

Organic matter content is inversely proportional to spectral reflectance. Although the organic matter content of the samples studied ranged from 0.09-9.0%, this parameter does not appear to mask out the contributions of other soil parameters to spectral variation. The iron content of the soil was significant in both visible and infrared regions. Its significance appeared to be unaffected by the presence of organic matter. The significance of iron increased with increasing wavelengths, possibly due to the vibrations and overtones of iron compounds in the middle and far infrared. Clay content was significant in the 0.50 to

Table 2.6-1 Physicochemical Properties of Soils Which Make a Significant Contribution in Explaining Variations in Visible and Infrared Reflectance.*

Spectral Interval (μm)	Variables** Entered	r-value
0.50-0.53	3,1,2	0.458
0.53-0.56	3,1,2	0.563
0.56-0.59	3,1	0.567
0.59-0.61	3,1	0.571
0.61-0.64	3,1,5	0.575
0.64-0.67	3,1,5	0.580
0.67-0.70	3,1,4	0.584
0.79-0.84	3,13,2	0.592
0.85-0.90	3,13,2	0.597
0.91-0.96	3,13,2	0.610
0.97-1.02	3,13,2	0.619
1.03-1.08	3,13,2	0.627
1.09-1.14	3,13,2	0.659
1.15-1.20	3,13,2	0.664
1.29-1.31	3,13,2	0.672
1.39-1.41	3,13,2	0.693
1.49-1.51	3,2,13	0.694
1.79-1.81	3,2,13	0.693
1.89-1.91	3,2,4	0.701
2.19-2.21	3,13,2	0.743
2.29-2.31	3,13,2	0.735
2.36-2.38	3,13,2	0.740

*Significant at 5% level

**Key to variables: 1 - Organic Matter, 2 - Iron, 3 - Silt, 4 - Clay, 5 - Color, 13 - CEC

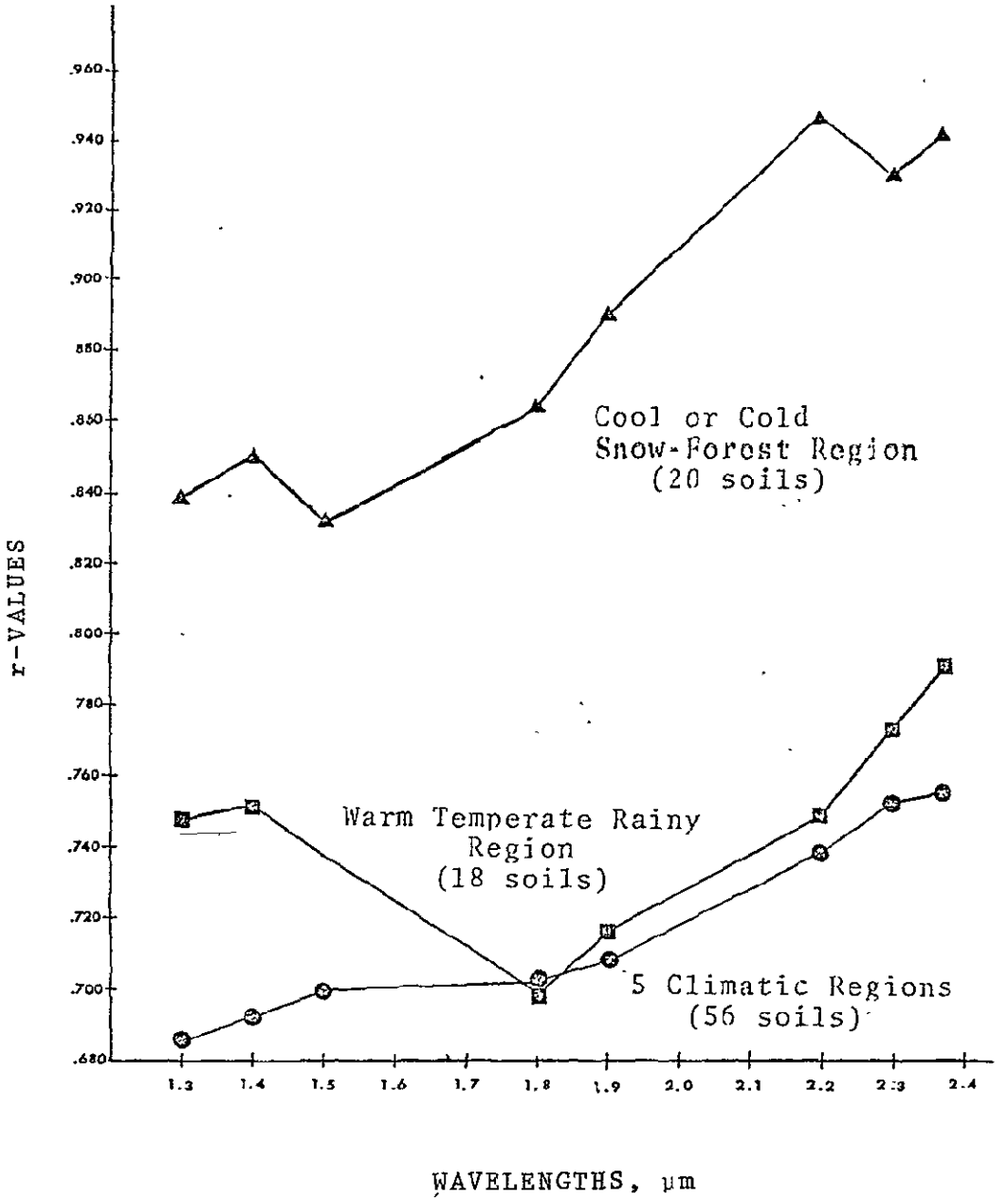


Figure 2.6-8 Climate effects on the correlation between infrared reflectance and some physico-chemical properites of soils.

0.70μm region and around the absorption bands at 1.4, 1.9 and 2.2μm. Increased clay content results in the attraction of larger quantities of atmospheric water to the soil surface, resulting in stronger absorption bands. The cation exchange capacity (CEC) was highly correlated (r=0.60) with spectral reflectance. CEC appears to be acting as an indicator for some of the natural interactions occurring between the soil properties. Organic matter/clay, organic matter/iron, and iron/silt interactions were significantly correlated with reflectance (r=0.65). In some instances these interactions were more significant in explaining spectral variation than the individual parameters. The mineralogy of the soils, to a large degree, determine the significance of several of the soil parameters (i.e., CEC) in explaining spectral variations in the visible and infrared regions. Cognizance of the dominant mineralogical composition of an area allows for the prediction of the relationship between spectral reflectance and soil properties.

The results of the statistical analysis of the relationship between biochemical properties and spectral reflectance suggests that there is a significant correlation between the two. Additional research is required before these relationships can be expressed quantitatively.

Results have shown that the grouping of soils from different climatic zones results in a lowering of the correlation between reflectance and the physicochemical properties of the soil. The effect of climate on the physicochemical nature of soils is at present a qualitative expression. The climatic factor cannot be satisfactorily quantified for use in these types of analysis. The best results were obtained when studying soils from within climatic regions (Figure 2.6-8).

The near and middle infrared regions offer excellent possibilities, in many respects superior to visible reflectance, for quantifying and defining the relationships between physicochemical properties of soils and spectral response (Table 2.6-1).

Although the effect of the physicochemical properties of soils on spectral reflectance is fully understood, the results presented here demonstrate the potential for developing methods of quantifying these relationships and sets the stage for continuing investigations in this area.

REFERENCES

1. Riemer, T.E., McGillem, C.D. Sept. 1974. Optimum Constrained Image Restoration Filters. Information Note 091974, Laboratory for Applications of Remote Sensing. See also SR&T Final Report for period June 1, 1974, May 31, 1975.

2. Hunt, R.M., Salisbury, J.W. 1970. Visible and Near-Infrared Spectra of Minerals and Rocks: I Silicate Minerals, Vol. 1.

ACKNOWLEDGEMENTS

LARS personnel contributing to the work described in this section are as follows: Ancillary Data Registration; Professional - Paul E. Anuta, Graduate - Nim-Chu and Martin Svedlow. The image enhancement evaluation work was done by graduate students Bijan Mobasseri and Dennis Adams and the soil properties work was done by graduate student Oscar Montgomery. Temporal Registration of LACIE site data was handled by the LARS computer facility with David Freeman the task leader.

Appendix C

Fortran Listings for IFOV Filter Program

```
FILE. . .     CONMAIN  FORTRAN  P1

 105  FORMAT(5X,20A4)                                                  CON03130
      WRITE(16, 110)                                                   CON03140
 110  FORMAT(/'ERROR ON RESULTS CARD. REENTER.')                       CON03150
      GO TO 15                                                         CON03160
 200  II = 1                                                           CON03170
      CALL IVAL(CARD, COL, NRUN, II, &100)                             CON03180
      FLAG(6) = .TRUE.                                                 CON03190
      GO TO 5                                                          CON03200
 300  II = 1                                                           CON03210
      CALL IVAL(CARD,COL, TAPENO, II, &100)                            CON03220
      FLAG(7) = .TRUE.                                                 CON03230
      GO TO 5                                                          CON03240
 400  II = 1                                                           CON03250
      CALL IVAL(CARD, COL, FILENO, II, &100)                           CON03260
      GO TO 5                                                          CON03270
      END                                                              CON03280
      SUBROUTINE BLKSUB(CARD, COL, FLAG, RUN, NFROW, NLROW,NFCOL,      CON03290
     1 NLCOL, *)                                                       CON03300
C                                                                      CON03310
C ***** INTERPRETS BLOCK CARD                                          CON03320
C                                                                      CON03330
      IMPLICIT INTEGER (A-Z)                                           CON03340
      LOGICAL * 1 FLAG(1)                                              CON03350
      DIMENSION PARLST(3), NUM(2), CARD(7)                             CON03360
      DATA PARLST/4HRUN ,4HLINE, 4HCOLS/                               CON03370
 1    FLAG(1) = .TRUE.                                                 CON03380
      IF(COL .EQ. 72) GO TO 100                                        CON03390
 5    IF(COL.EQ.72) GO TO 10                                           CON03400
      CALL CTLPRM(CARD, COL, PARLST, 3, CODE, &15)                     CON03410
      GO TO (200, 300, 400), CODE                                      CON03420
 10   IF(FLAG(2)) RETURN 1                                             CON03430
 15   COL = 0                                                          CON03440
      ERR = 1                                                          CON03450
      CALL CTLWRD(CARD, COL, 4HBLOC, 1, CODE, 5, ERR)                  CON03460
      IF(ERR .EQ. 4) STOP                                              CON03470
      GO TO 5                                                          CON03480
 100   WRITE(16,105) CARD                                              CON03490
 105   FORMAT(5X,20A4)                                                 CON03500
      WRITE(16,110)                                                    CON03510
 110  FORMAT(/'ERROR ON BLOCK CARD. REENTER.')                         CON03520
      GO TO 15                                                         CON03530
 200  II = 1                                                           CON03540
      CALL IVAL(CARD,COL,RUN,II, &100)                                 CON03550
      FLAG(2) = .TRUE.                                                 CON03560
      GO TO 5                                                          CON03570
 300  I2 = 2                                                           CON03580
      CALL IVAL(CARD, COL, NUM, I2, &150)                              CON03590
      NFROW = NUM(1)                                                   CON03600
      NLROW = NUM(2)                                                   CON03610
      FLAG(3) = .TRUE.                                                 CON03620
      GO TO 5                                                          CON03630
 400  I2 = 2                                                           CON03640
      CALL IVAL (CARD, COL, NUM, I2, &107)                             CON03650
      NFCOL = NUM(1)                                                   CON03660
      NLCOL = NUM(2)                                                   CON03670
      FLAG(4) = .TRUE.                                                 CON03680
      GO TO 5                                                          CON03690
      END                                                              CON03700
      SUBROUTINE CHISUB(CARD, COL, FLAG, ICHAN, NCHAN, *)              CON03710
C                                                                      CON03720
C ***** INTERPRETS CHANNEL CARDS                                       CON03730
C                                                                      CON03740
      IMPLICIT INTEGER (A-Z)                                           CON03750
      LOGICAL * 1 FLAG(1)                                              CON03760
      INTEGER * 2 ICHAN,  CSEL(3)), CHAN(3))                           CON03770
      REAL CSET(90)                                                    CON03780
      DIMENSION ICHAN(30), CARD(7)                                     CON03790
      DATA TOP/1/, CSEL/30*0/,CSET/30*-9000./,CHAN/30*0/               CON03800
 5    CALL CHANEL(CARD, COL, NCP, CSEL, CSET, CHAN, &107)              CON03810
      DO 10 I = TOP, NCP                                               CON03820
      K = CHAN(I)                                                      CON03830
      IF(ICHAN(K) .EQ. 1 ) GO TO 10                                    CON03840
      ICHAN(K) = 1                                                     CON03850
      NCHAN = NCHAN + 1                                                CON03860
 10   CONTINUE                                                         CON03870
      TOP = NCP + 1                                                    CON03880
      FLAG(8) = .TRUE.                                                 CON03890
      RETURN 1                                                         CON03900
```

```
FILE. . .     CCANAIN  FORTRAN  P1

 100  WRITE(16, 105) CARD                                              CON03910
 105  FORMAT(5X,20A4)                                                  CON03920
      WRITE(16,110)                                                    CON03930
 110  FORMAT(/'ERROR ON CHANNELS CARD. REENTER.')                      CON03940
      COL = 0                                                          CON03950
      ERR = 1                                                          CON03960
      CALL CTLWRD(CARD,COL, 4HCHAN, 1, CODE, 5, ERR)                   CON03970
      IF(ERR .EQ. 4) STOP                                              CON03980
      GO TO 5                                                          CON03990
      END                                                              CON04000
```

```
FILE: , ,     CONVO    FORTRAN  P1

      SUBROUTINE CONVL(CYLIN, IOBUF, NFCOL, NLCOL, NFROW, NLROW, NPCOL, CON00010
     1 NPROW, CUTFF, CUFFV, NYCOL, NYROW, [NTAPE, NSKIP, NSAMP, ICHAN, CON00020
     2 INCHN,NCHAN, ALLCHN, CAL, NESOUT, IEXTLV)                       CON00030
      REAL*4 X(1000), CUTFF(1), CUFFV(1)                               CON00040
      LOGICAL*1 CYLIN(1), IOBUF(1), CAL(1), N1(2), N2(2),              CON00050
     1 IEXTLV, ALLCHN                                                  CON00060
      INTEGER*2 I1/0/, I2/0/,  ICHANE(1)                               CON00070
      DIMENSION POINTER(5)                                             CON00080
      EQUIVALENCE (N1,I1), (N2,I2)                                     CON00090
C                                                                      CON00100
C  ***** SET UP CONSTANTS                                             CON00110
C                                                                      CON00120
      NP1 = NNROW + 1                                                  CON00130
      KOUNT = 1                                                        CON00140
      KOUNTM = NLROW - NFROW + 1                                       CON00150
      LREC = [NSAMP + 6]*INCHN + 4                                     CON00160
      NPCOL = 2 * NPCOL + 1                                            CON00170
      NPROW = 2 * NPROW + 1                                            CON00180
      ICYLLN = [NLCOL - NFCOL + 1] * [NNCOL +1]                        CON00190
      ICYROW = ICYLLN * NCHAN                                          CON00200
      IEXTLV = .FALSE.                                                 CON00210
      NBUF = ICYROW * NCHAN * 6 + 4                                    CON00220
      IPCT = 10                                                        CON00230
C                                                                      CON00240
C  ***** INITIAIL CYLINDER                                            CON00250
C                                                                      CON00260
  10  IOBASE = 0                                                       CON00270
      DO 200 ICY = 1, INPROW                                           CON00280
      INBASE = 4                                                       CON00290
C                                                                      CON00300
C  ***** READ A LINE AND ISOLATE SELECTED CHANNELS                    CON00310
C                                                                      CON00320
      CALL TOPKD(INTAPE, LREC, ERR, IOBUF)                             CON00330
      IF(ERR .EQ. 0) GO TO 15                                          CON00340
  11  WRITE(ERROUT, 12) ERR                                            CON00350
  12  FORMAT('ERROR', I5, ' IN TOPKD.')                                CON00360
      IEXTLV = .TRUE.                                                  CON00370
      RETURN                                                           CON00380
  15  IF(ALLCHN) GO TO 21                                              CON00390
      IKNT = 0                                                         CON00400
      JJ = 4                                                           CON00410
      NB = NSAMP + 6                                                   CON00420
      DO 16 I = 1, 32                                                  CON00430
      IF(ICHAN(I) .EQ. 0) GO TO 16                                     CON00440
      IKNT = IKNT + 1                                                  CON00450
      IBASE = 4 + (I - 1) * NB                                         CON00460
      IF(IBASE.NE.JJ) GO TO 22                                         CON00470
      JJ = JJ + NB                                                     CON00480
      GO TO 23                                                         CON00490
  22  CALL MOVBYT(IOBUF, IBASE, 1, IOBUF, JJ, 1, NB)                   CON00500
      JJ = JJ + NB                                                     CON00510
  23  IF(IKNT.EQ. NCHAN) GO TO 21                                      CON00520
  16  CONTINUE                                                         CON00530
C                                                                      CON00540
C  ***** INTERPOLATE HORIZONTALLY FROM THE BUFFER TO THE CYLINDER      CON00550
C                                                                      CON00560
  21  DO 100 ICHN = 1, NCHAN                                           CON00570
      DO 30 ICOL = NFCOL, NLCOL                                        CON00580
      JJ = IOBASE + [ICOL - NFCOL] * [NNCOL + 1] + 1                   CON00590
      IRC = ICOL - NFCOL + 1 + INBASE                                  CON00600
      X = JJ                                                           CON00610
      DO 17 J = 1, IAPCOL                                              CON00620
      I2 = 0                                                           CON00630
      N2(2) = IOBUF(INC + J)                                           CON00640
      X = X + CUEFF(IJ) * I2                                           CON00650
  17  CONTINUE                                                         CON00660
      II = X                                                           CON00670
      CYLIN(JJ) = N1(2)                                                CON00680
      DO 29 KK = 1, NYCOL                                              CON00690
      KJ = KK * INPCOL                                                 CON00700
      X = 0.0                                                          CON00710
      DO 18 J = 1, INPCOL                                              CON00720
      I2 = 0                                                           CON00730
      N2(2) = IOBUF(INC + J)                                           CON00740
      X = X + CUEFF(KJ + J) * I2                                       CON00750
  18  CONTINUE                                                         CON00760
      II = X                                                           CON00770
      CYLIN(KK + JJ) = N1(2)                                           CON00780
```

```
FILE: , ,     CONVO    FORTRAN  P1

  20  CONTINUE                                                         CON00790
  30  CONTINUE                                                         CON00800
C                                                                      CON00810
C  ***** SAVE CALIBRATIONS                                            CON00820
C                                                                      CON00830
      INBASE = INBASE + NSAMP                                          CON00840
      KL = [[IL] - 1] * NCHAN + [ICHN - 1] * 6                         CON00850
      DO 40 I = 1, 6                                                   CON00860
      CAL(KL + I) = IOBUF(INBASE + I)                                  CON00870
  40  CONTINUE                                                         CON00880
      INBASE = INBASE + 6                                              CON00890
      ICBASE = ICBASE + ICYLLN                                         CON00900
 100  CONTINUE                                                         CON00910
 200  CONTINUE                                                         CON00920
      ITOP = 1                                                         CON00930
      NIC = NPROW + 1                                                  CON00940
      LINE = 1                                                         CON00950
      IACT = INPROW                                                    CON00960
C                                                                      CON00970
C  ***** SET UP CYLINDER POINTERS                                     CON00980
C                                                                      CON00990
      IBASE = 0                                                        CON01000
      DO 210 I = 1, INPROW                                             CON01010
      POINTER(I) = IBASE                                               CON01020
      IBASE = IBASE + ICYROW                                           CON01030
 210  CONTINUE                                                         CON01040
C                                                                      CON01050
C  ***** PUT CALIBRATION AND LINE NUMBERS IN OUTPUT BUFFER            CON01060
C                                                                      CON01070
 211  IBASE = 4                                                        CON01080
      DO 250 I = 1, NP1                                                CON01090
      ICB = [NIC - 1] * NCHAN * 6                                      CON01100
      I1 = LINE                                                        CON01110
      IOBUF(IBASE - 3) = N1(1)                                         CON01120
      IOBUF(IBASE - 2) = N1(2)                                         CON01130
      I1 = ITOP                                                        CON01140
      IOBUF(IBASE - 1) = N1(1)                                         CON01150
      IOBUF(IBASE) = N1(2)                                             CON01160
      ICBS = IBASE + ICYLLN                                            CON01170
      DO 220 K = 1, NCHAN                                              CON01180
      DO 215 IK = 1, 6                                                 CON01190
      IOBUF(ICBS + IK) = CAL(ICB + IK)                                 CON01200
 215  CONTINUE                                                         CON01210
      ICBS = ICBS + ICYLLN + 6                                         CON01220
      ICB = ICB + 6                                                    CON01230
 220  CONTINUE                                                         CON01240
      LINE = LINE + 1                                                  CON01250
      IBASE = IBASE + NBUF                                             CON01260
 250  CONTINUE                                                         CON01270
C                                                                      CON01280
C  ***** INTERPOLATE VERTICALLY INTO OUTPUT BUFFER                    CON01290
C                                                                      CON01300
      IBASE = 4                                                        CON01310
      ICHN = 0                                                         CON01320
      DO 340 I =1, NCHAN                                               CON01330
      DO 330 J = 1, ICYLLN                                             CON01340
      DO 299 M = 1,NP1                                                 CON01350
      XX(M) = 0.0                                                      CON01360
 299  CONTINUE                                                         CON01370
      DO 322 K = 1, INPROW                                             CON01380
      IP = POINTER(K)                                                  CON01390
      II = 0                                                           CON01400
      N1(2) = CYLIN(IP + ICHN + J)                                     CON01410
      X = II                                                           CON01420
      ITMP = K                                                         CON01430
      DO 320 L = 1, NP1                                                CON01440
      XX(L) = XX(L) + CUEFV(ITMP) * X                                  CON01450
      ITMP = ITMP + INPROW                                             CON01460
 320  CONTINUE                                                         CON01470
 320  CONTINUE                                                         CON01480
      KK = IBASE + J                                                   CON01490
      DO 325 L = 1, NP1                                                CON01500
      II = XX(L)                                                       CON01510
      IOBUF(KK) = N1(2)                                                CON01520
      KK = KK + NBUF                                                   CON01530
 325  CONTINUE                                                         CON01540
 330  CONTINUE                                                         CON01550
      IBASE = IBASE + ICYLLN + 6                                       CON01560
```

```
FILE: , ,     CONMAIN  FORTRAN  P1

      KNT = 0                                                          CON01570
      JJ = 31                                                          CON01580
      DO 270 I = 1, 105                                                CON01590
      IF(ICHANE(I) .EQ. 0) GO TO 270                                   CON01600
      KNT = KNT + 1                                                    CON01610
      IBASE = 31 + (I - 1) * 5                                         CON01620
      ITOP = IBASE + 4                                                 CON01630
      DO 269 J = IBASE, ITOP                                           CON01640
      RIDE(JJ) = N1DE(J)                                               CON01650
      JJ = JJ + 1                                                      CON01660
 269  CONTINUE                                                         CON01670
      IF(KNT .EQ. NCHAN) GO TO 283                                     CON01680
 270  CONTINUE                                                         CON01690
      WRITE(ERROUT, 275)                                               CON01700
 275  FORMAT('ERROR0 ILLEGAL CHANNEL NUMBER SPECIFIED')                CON01710
      GO TO 299                                                        CON01720
 280  DO 290 I = JJ, 999                                               CON01730
      RIDE(I) = 0.0                                                    CON01740
 290  CONTINUE                                                         CON01750
C                                                                      CON01760
C  ***** TYPE OUT PARAMETERS                                          CON01770
C                                                                      CON01780
 300  WRITE(CONOUT, 310) KOUNT,NFCOL, NLCOL, NFROW, NLROW, NROW,TAPENO, CON01790
     1 FILENO                                                          CON01800
 310  FORMAT('////**** INPUT ****//'ROW NUMBER ', I2,'[FIRST COLUMN ', I2,'CON01810
     1/'LAST COLUMN ', I2,'/'[ESS]'LINE 0 ', I2/'LAST LINE 0 ', I2/     CON01820
     1'**** OUTPUT ****//'ROW NUMBER 0 ', I2/'TAPE NUMBER ', I2/'FILE NCON01830
     1UMBER ', I4)                                                     CON01840
      IF(.NOT. ALLCHN) GO TO 314                                       CON01850
      WRITE(CONOUT, 311)                                               CON01860
 317  FORMAT('//'**** ALL INPUT CHANNELS SELECTED ****')               CON01870
      GO TO 316                                                        CON01880
 314  WRITE(CONOUT, 311)                                               CON01890
 311  FORMAT('//'**** CHANNEL ****//'    IN  OUT')                    CON01900
      K = 1                                                            CON01910
      DO 313 I = 1, 32                                                 CON01920
      IF(ICHAN(I) .EQ. 1) WRITE(CONOUT,312) I,K                        CON01930
 312  FORMAT(2I6)                                                      CON01940
      K = K + 1                                                        CON01950
 313  CONTINUE                                                         CON01960
C                                                                      CON01970
C  ***** MOUNT OUTPUT TAPE AND WRITE ID                               CON01980
C                                                                      CON01990
 318  CALL MOUNT(TAPENO, 12, *911)                                     CON02000
      IF(FILENO .EQ. 1) GO TO 424                                      CON02010
      I = FILENO - 1                                                   CON02020
      DO 313 J = 1, I                                                  CON02030
      CALL TOPFF(12)                                                   CON02040
 315  CONTINUE                                                         CON02050
 320  CNT = 800                                                        CON02060
      CALL TOPWR(12, CNT, ERR, IO)                                     CON02070
      IF(ERR .EQ. 0) GO TO 360                                         CON02080
      WRITE(CONOUT, 321) ERR                                           CON02090
 330  FORMAT('ERROR NUMBER ', I5, ' IN TOPWR.')                        CON02100
      GO TO 1000                                                       CON02110
C                                                                      CON02120
C  ***** SKIP TO FIRST LINE OF INPUT                                  CON02130
C                                                                      CON02140
 340  J = NFROW - NPROW - 1                                            CON02150
      IF(J .EQ. 0) GO TO 360                                           CON02160
      CALL TOFES(11, J, ERR)                                           CON02170
      IF(ERR .EQ. 0) GO TO 360                                         CON02180
      WRITE(CONOUT, 350) ERR                                           CON02190
 350  FORMAT('ERROR ',I3,' IN TOFES.')                                 CON02200
      GO TO 1000                                                       CON02210
C                                                                      CON02220
C  ***** GENERATE COEF. MATRIX                                        CON02230
C                                                                      CON02240
 360  DO 370 I = 1, 6                                                 CON02250
      N1(I) = N1(I+4) = I                                              CON02260
 370  CONTINUE                                                         CON02270
      DO 380 I = 1,66                                                  CON02280
      N2(I) = N2(I+6) = I                                              CON02290
 380  CONTINUE                                                         CON02300
      DO 390 J = 1, 2                                                  CON02310
      COEFM(J, 1) = N1(IJ-1)* 12 + 1)                                  CON02320
 390  CONTINUE                                                         CON02330
      K = 9                                                            CON02340
```

```
FILE: , ,     CONMAIN  FORTRAN  P1

      DO 400 L = 2, 3                                                  CON02350
      COEFM(1, L) = 1.0                                                CON02360
      DO 395 J = 2,9                                                   CON02370
      COEFM(J,L) = N1(IJ - 2) * 12 + K)                                CON02380
 395  CONTINUE                                                         CON02390
      K = 5                                                            CON02400
 400  CONTINUE                                                         CON02410
      DO 410 J = 1,9                                                   CON02420
      COEFV(J, 1) = N2(IJ-1)* 16 + 1)                                  CON02430
 410  CONTINUE                                                         CON02440
      K = 13                                                           CON02450
      DO 430 L = 2, 6                                                  CON02460
      COEFV(1, L) = 0.0                                                CON02470
      DO 420 J = 2,9                                                   CON02480
      COEFV(J,L) = N2(IJ-2) * 16 + K)                                  CON02490
 420  CONTINUE                                                         CON02500
      K = K - 4                                                        CON02510
 430  CONTINUE                                                         CON02520
C                                                                      CON02530
C  ***** PERFORM CONVOLUTION                                          CON02540
C                                                                      CON02550
      CALL CONVL(CYLIN, IOBUF, NFCOL, NLCOL, NFROW, NLROW, NPCOL, NPROW,CON02560
     1 COEFM, COEFV, NYCOL, NYROW, 11, 12, [106, ICHAN,155, NCHAN,ALLCH CON02570
     2 CAL, IOMOUT, IEXTLV)                                            CON02580
C                                                                      CON02590
C  ***** CHECK FOR ERRORS                                             CON02600
C                                                                      CON02610
      IF(.NOT.IEXTLV) GO TO 470                                        CON02620
      GO TO 1000                                                       CON02630
C                                                                      CON02640
C  ***** WRITE ELF RECORD                                             CON02650
C                                                                      CON02660
 470,     ID(3) = J                                                    CON02670
      ID(1) = NFLD                                                     CON02680
      ID(2) = FILENO                                                   CON02690
      ID(4) = J                                                        CON02700
      DO 480 J = 5, 52                                                 CON02710
      ID(J) = 0                                                        CON02720
 480  CONTINUE                                                         CON02730
      DO 490 J = 31, 203                                               CON02740
      RIDE(J) = 0.0                                                    CON02750
 490  CONTINUE                                                         CON02760
      KNT = 200                                                        CON02770
      CALL TOPLE(12)                                                   CON02780
      CALL TOPWR(12, KNT, ERR, IO)                                     CON02790
      IF(ERR .EQ. 0) GO TO 515                                         CON02800
      WRITE(CONOUT, 501) ERR                                           CON02810
 500  FORMAT('//'ERROR NUMBER ', I3, ' IN TOP WR.')                    CON02820
      GO TO 1000                                                       CON02830
 510  CALL TOPLE(12)                                                   CON02840
      CALL TOPEF(12)                                                   CON02850
1000 CALL TOPRW(11).                                                   CON02860
      CALL TOPWU(12)                                                   CON02870
      STOP                                                             CON02880
      END                                                              CON02890
      SUBROUTINE RESSUB(CARD,COL, FLAG, NROW, TAPENO, FILENO, *)       CON02900
C                                                                      CON02910
C  ***** INTERPRETS RESULTS CARDS                                     CON02920
C                                                                      CON02930
      IMPLICIT INTEGER (A-Z)                                           CON02940
      LOGICAL*1 FLAG(1)                                                CON02950
      DIMENSION PARLST(4), CARD(20)                                    CON02960
      DATA PARLST/'NROW', 'TAPE', 'FILE'/                              CON02970
      FLAG(5) = .TRUE.                                                 CON02980
C                                                                      CON02990
C  ***** SEARCH FOR PARAMETER                                         CON03000
C                                                                      CON03010
      IF(COL .LE. 72) GO TO 100                                        CON03020
  5   IF(COL .LE. 72) GO TO 15                                         CON03030
      CALL CTLPRP(CARD, COL, PARLST, 3, CODE, *100)                    CON03040
      GO TO (200, 100, 400), CODE                                      CON03050
 10   IF(FLAG(6).AND.FLAG(7)) RETURN 1                                 CON03060
 15   ERR = 1                                                          CON03070
      COL = 0                                                          CON03080
      CALL CTLWRD(CARD, COL, 4HRESU, 1, CODE, 5, ERR)                  CON03090
      IF(ERR .LE. 4) STOP                                              CON03100
      GO TO 5                                                          CON03110
 100  WRITE(16,105) CARD                                               CON03120
```

FILE. . . CONMAIN FORTRAN P1

```
C                                                                       CON00010
C *****  MAIN PROGRAM FOR CONVOLUTION SYSTEM                            CON00020
C *****                    DENNIS C. ADAMS                              CON00030
C                                                                       CON00040
      IMPLICIT INTEGER (A-Z)                                            CON00050
      LOGICAL*1 FLAG(8), ALLCHN, IER                                    CON00060
      INTEGER * 2 ICHAN                                                 CON00070
      REAL KID, HI, HZ, CHERN, CONFV                                    CON00080
      DIMENSION CELLSIZE(6), CARD(20), ICHANI(6), ID(200), KID(200), HI(97CON00090
     1),HZ(1297, COLFW(97), COLFV(97,4), CAL(324), ICBUF(1400), CYLIN  CON00100
     2(250000)                                                          CON00110
      DATA HI/ 8.0005316E-04, 1.0229760E-03, 1.3286113E-03              CON00120
     +,  1.4944201E-03, 1.3744208E-03, 9.8671495E-04                    CON00130
     +,  4.2271236E-04,-4.8972791E-04,-1.6234147E-03                    CON00140
     +, -3.0933078E-03,-4.4490171E-03,-5.6214952E-03                    CON00150
     +, -5.9452519E-03,-5.4437376E-03,-3.5124416E-03                    CON00160
     +, -1.0357E-03,-8.8894459E-03,-6.6345478E-03                       CON00170
     +, -4.8724704E-04, 5.8342679E-03, 2.4428368E-02                    CON00180
     +,  4.6900189E-02, 5.1011746E-02, 5.4466211E-02                    CON00190
     +,  4.7416626E-02, 3.9459694E-02, 1.9774113E-02                    CON00200
     +, -5.4434772E-04,-2.1143741E-02,-4.2153022E-02                    CON00210
     +, -5.1326268E-02,-5.3933292E-02,-2.2918934E-  1                   CON00220
     +, -3.1007263E-01,-3.1211269E-01,-3.0136461E-01                    CON00230
     +, -2.4891251E-01,-1.8294481E-01,-8.9776893E-02                    CON00240
     +,  5.2803L6E-02, 2.5276.55E-01, 4.5393956E-01                     CON00250
     +,  6.1992.78E-01, 7.4028177E-01, 9.3741135E-01                    CON00260
     +,  1.1272911 00, 1.3242311E 00, 1.4068346E 00                     CON00270
     +,  1.4171120E 00, 1.49 .6.E/                                      CON00280
      DATA HZ/-7.8736735E-04,-7.8485780E-04,-7.6378419E-04              CON00290
     +, -6.9729.2E-04,-5.1114.91E-04,-2.2120363E-04                     CON00300
     +,  5.9114019E-04, 2.0643714E-04,-1.0884942E-04                    CON00310
     +,  2.3196277E-04, 1.6648.34E-04, 2.5133739E-04                    CON00320
     +,  3.1933.64E-04, 2.6766552E-04, 1.7127.56E-04                    CON00330
     +, -1.0594130E-04,-4.4613242E-04,-1.0130906E-03                    CON00340
     +, -1.6278633E-03,-2.0931645E-03,-2.2768580E-03                    CON00350
     +, -2.2863761E-03,-2.0933492E-03,-1.7005334E-03                    CON00360
     +, -1.2965279E-03,-8.5272396E-04,-3.1571249E-04                    CON00370
     +, -6.1868472E-04, 1.2331429E-03, 2.4816995E-02                    CON00380
     +,  2.5847.48E-02, 2.6855955E-02, 2.7424808E-02                    CON00390
     +,  1.0260477E-02, 2.7671295E-  3,-6.1644938E-03                   CON00400
     +, -2.5945.28E-02,-4.0196527E-02,-5.1822.54E-02                    CON00410
     +, -7.1651.)3E-02,-8.1331005E-02,-8.8932471E-02                    CON00420
     +, -1.0240295E-01,-1.1352876E-01,-1.2242075E-01                    CON00430
     +, -1.2804.51E-01,-1.2.12.-.59E-01,-1.2469723E-01, -1.4850965E-01  CON00440
     +, -1.0277433E-01,-7.7431154E-02, 5.1625595E-04, 1.8402124E-01     CON00450
     +,  3.1243.2E-01, 4.  .1277E-01, 4.3732432E-01, 5.0.7.0.2E-01      CON00460
     +,  6.7320.2E-01, 7.6664666E-01, 8.5631.41E-01, 9.5.01277E-01      CON00470
     +,  1.13 .4.E 00, 1.2.7.737E 00, 1.2913467, 1.2.60747, 1.6.9.2/    CON00480
      DATA FLAG/8*.FALSE./, CELLSIZE/.CELLSIZE./, IPASS/ .TRUE./,       CON00490
     1 FILEND/7*.  /, ICHAN/33*2/, CONOUT/16/, ALLCHN/.FALSE./,         CON00500
     2 NCHAN, NCOL, NPROW, NPCOL/ 4*1/, NCHAN/6/                        CON00510
      EQUIVALENCE (ID, I7), (CAL(1), HZ), (CAL, HI), (ICBUF, ID)        CON00520
C                                                                       CON00530
C *****  INTERPRET CONFROM CARD DECK                                    CON00540
C *****  READ A CONTROL CARD AND ESTABLISH TYPE                         CON00550
C                                                                       CON00560
   10 ERR = 0                                                           CON00570
      COL = 0                                                           CON00580
      CALL GTLAMC(CARD, COL, CELLSI, 4, CODE, 5, ERR)                   CON00590
      IF(CODE .EQ. 4) STOP                                              CON00600
      IF(ERR .EQ. 2) GO TO 130                                          CON00610
      GO TO (100, 110, 120, 130), CODE                                  CON00620
  100 CALL BLRSCR(CARD, COL, FLAG, 8, NERROR, ALARM, NFCOL, NLCOL, 61 ICON00630
  110 CALL MESSAGE(CARD, COL, FLAG, NAME, TAPENO, FILENO, &13)          CON00640
  120 CALL CHANSP(CARD, COL, FLAG, ICHAN, NCHAN, &13)                   CON00650
C                                                                       CON00660
C *****  END OF CONTROL CARD FILE OR END CARD.                          CON00670
C *****  CHECK FOR BLOCK CARD                                           CON00680
C                                                                       CON00690
  130 IF(FLAG(11)) GO TO 150                                            CON00700
      WRITE(CONOUT, 140)                                                CON00710
  140 FORMAT(/'BLOCK CARD REQUIRED. PLEASE ENTER.')                     CON00720
      COL = 0                                                           CON00730
      ERR = 3                                                           CON00740
      CALL GTLAMC(CARD, COL, CELLSI(1), 1, CODE, 5, ERR)                CON00750
      IF(ERR .EQ. 4) STOP                                               CON00760
      CALL BLRSCR(CARD, COL, FLAG, 8, NERROR, ALARM, NFCOL, NLCOL, &15)CON00770
1     1 )                                                               CON00780
```

FILE. . . CONMAIN FORTRAN P2

```
C                                                                       CON00790
C *****  CHECK FOR RESULTS CARD                                         CON00800
C                                                                       CON00810
  150 IF(FLAG(5)) GO TO 17.                                             CON00820
      WRITE(CONOUT, 160)                                                CON00830
  160 FORMAT(/'RESULTS CARD REQUIRED. PLEASE ENTER.')                   CON00840
      COL = 0                                                           CON00850
      ERR = 3                                                           CON00860
      CALL GTLAMC(CARD, COL, CELLSI(2), 1, CODE, 5, ERR)                CON00870
      IF(ERR .EQ. 4) STOP                                               CON00880
      CALL MESSAGE(CARD, COL, FLAG, NAME, TAPENO, FILENO, &170)         CON00890
C                                                                       CON00900
C *****  CHECK FOR CHANNEL DEFAULTS                                     CON00910
C                                                                       CON00920
  170 IF(FLAG(4)) GO TO 190                                             CON00930
      DO 180 J = 1, 3                                                   CON00940
      ICHAN(J) = J                                                      CON00950
  180 CONTINUE                                                          CON00960
C                                                                       CON00970
C *****  MOUNT THE INPUT TAPE AND CONSTRUCT THE ID RECORD FOR NEW TAPE  CON00980
C                                                                       CON00990
  190 ID(1) = 0                                                         CON01000
      CALL GADMOUNT(NUM, 1), 10, ERR)                                   CON01010
      IF(ERR .EQ. 0) GO TO 210                                          CON01020
      WRITE(CONOUT, 200) ERR, NUM                                       CON01030
  200 FORMAT('ERROR NUMBER ', I3, ' IN MOUNTING RUN NUMBER ', I8,'.')   CON01040
  205 CALL TOPMOF(1)                                                    CON01050
      STOP                                                              CON01060
  210 ID(1) = TAPEND                                                    CON01070
      ID(2) = FILEND                                                    CON01080
      ID(3) = NRUN                                                      CON01090
      ID5 = ID(5)                                                       CON01100
      IF(.NOT. FLAG(8)) NCHAN = ID(5)                                   CON01110
      IF(NCHAN.LE.ID(5)) GO TO 240                                      CON01120
      WRITE(CONOUT, 230) NCHAN, ID(5)                                   CON01130
  230 FORMAT('ERROR ',I3,' CHANNELS SELECTED  ONLY ',I3,' EXIST.')      CON01140
      GO TO 20.                                                         CON01150
  240 ID(5) = NCHAN                                                     CON01160
      IF(NCHAN .EQ. ID5) ALLCHN = .TRUE.                                CON01170
      IF(.NOT. FLAG(4)) GO TO 241                                       CON01180
      IF(NFCOL .NE.0 .OR. I) GO TO 243                                  CON01190
      NFCOL = 1 + NPCOL                                                 CON01200
      WRITE(CONOUT, 241) NFCOL, NFCOLI                                  CON01210
  241 FORMAT(/'FIRST COL. = ', I4, ' TOO SMALL  FIRST COL. = ',         CON01220
     1 I4, ' ASSUMED.')                                                 CON01230
      NFCOL = NFCOLI                                                    CON01240
  243 IF(NLCOL + NPCOL .LE. ID(6) - 6) GO TO 245                        CON01250
      NLCOL. = ID(6) -NPCOL - 6                                         CON01260
      WRITE(CONOUT, 244) NLCOL, NLCOLI                                  CON01270
  244 FORMAT(/'LAST COL. = ', I4, ' TOO LARGE  LAST COL. = ',           CON01280
     1 I4, ' ASSUMED.')                                                 CON01290
      NLCOL = NLCOLI                                                    CON01300
      GO TO 245                                                         CON01310
  251 NFCOL = 1 +.NPCOL                                                 CON01320
      NLCOL = ID(6) - 6 - NPCOL                                         CON01330
  245 NSAMP = (NLCOL-NFCOL+1)*(NNCOL+1)                                 CON01340
      ID6 = ID(6) - 6                                                   CON01350
      ID(6) = NSAMP + 6                                                 CON01360
      CALL GIDATE(ID(17))                                               CON01370
      IF(.NOT. FLAG(3)) GO TO 252                                       CON01380
      IF(NROW + NPROW .GE. 1) GO TO 247                                 CON01390
      NFROW1 = 1 + NPROW                                                CON01400
      WRITE(CONOUT, 246) NFROW, NFROW1                                  CON01410
  246 FORMAT(/'FIRST LINE = ', I4, ' TOO SMALL  FIRST LINE = ', I4,     CON01420
     1 ' ASSUMED.')                                                     CON01430
      NFROW = NFROW1                                                    CON01440
  247 IF(NLROW + NPROW .LE. ID(20)) GO TO 250                           CON01450
      NLROW1 = ID(20) - NPROW                                           CON01460
      WRITE(CONOUT, 248) NLROW, NLROW1                                  CON01470
  248 FORMAT(/'LAST LINE = ', I4, ' TOO LARGE  LAST LINE = ', I4,       CON01480
     1 ' ASSUMED.')                                                     CON01490
      NLROW = NLROW1                                                    CON01500
      GO TO 250                                                         CON01510
  252 NFROW = 1 +NPROW                                                  CON01520
      NLROW = ID(20) - NPROW                                            CON01530
  250 ID(20) = (NLROW - NFROW + 1) * (NNROW + 1)                        CON01540
      IF(ALLCHN) GO TO 300                                              CON01550
      IFIALLCHN) GO TO 300                                              CON01560
```

FILE. . . CONVO FORTRAN P1

```
      ICHN = ICHN + ICYLLN                                              CON01570
  340 CONTINUE                                                          CON01580
C                                                                       CON01590
C *****  WRITE OUTPUT RECORDS                                           CON01600
C                                                                       CON01610
      IBASE = 1                                                         CON01620
      DO 331 J = 1, NPI                                                 CON01630
      CALL TOPWR(NEWTAP, NBUF, IERR, IOBUF(IBASE))                      CON01640
      IF(IERR .NE. 0) GO TO 341                                         CON01650
      IBASE = IBASE + NBUF                                              CON01660
  331 CONTINUE                                                          CON01670
      GO TO 350                                                         CON01680
  341 WRITE(MESOUT, 345) IERR                                           CON01690
  345 FORMAT(/'ERROR', I3,' IN TOPWR.')                                 CON01700
      IERFIG = .TRUE.                                                   CON01710
      RETURN                                                            CON01720
  350 IPCT = IPCT -1                                                    CON01730
      IF(IPCT .NE. 0) GO TO 351                                         CON01740
      WRITE(MESOUT, 3521 KOUNT                                          CON01750
  352 FORMAT(I3,' LINES CONVOLUTED.')                                   CON01760
      IPCT = 5                                                          CON01770
  351 IF(KOUNT .GE. KOUNTM) RETURN                                      CON01780
      KOUNT = KOUNT + 1                                                 CON01790
C                                                                       CON01800
C *****  ROTATE POINTERS                                                CON01810
C                                                                       CON01820
      J = INPROW - 1                                                    CON01830
      K = POINTI(1)                                                     CON01840
      DO 360 I = 1, J                                                   CON01850
      POINTI(I) = POINTI(I + 1)                                         CON01860
  360 CONTINUE                                                          CON01870
      POINTI(INPROW) = K                                                CON01880
      ITOP = ITOP + NPROW                                               CON01890
      IF(ITOP .GT. INPROW) ITOP = ITOP - INPROW                         CON01900
      MID = ITOP + NPROW                                                CON01910
      IF(MID .GT. INPROW) MID = MID - INPROW                            CON01920
      IBOT = MID + NPROW                                                CON01930
      IF(IBOT .GT. INPROW) IBOT = IBOT - INPROW                         CON01940
C                                                                       CON01950
C *****  READ IN A NEW RECORD AND ISOLATE SELECTED CHANNELS             CON01960
C                                                                       CON01970
      ICBASE = POINTI(INPROW)                                           CON01980
      INBASE = 4                                                        CON01990
      CALL TOPRD(INTAPE, LNFG, IERR, IOBUF)                             CON02000
      IF(IERR .NE. 0) GO TO 11                                          CON02010
      IF(ALLCHN) GO TO 421                                              CON02020
      JJ = 4                                                            CON02030
      IKNT = 0                                                          CON02040
      DO 416 J = 1, 30                                                  CON02050
      IF(ICHAN(J) .EQ. 0) GO TO 416                                     CON02060
      IKNT = IKNT + 1                                                   CON02070
      IBASE = 4 + (I - 1) * NB                                          CON02080
      IF(IBASE .NE. JJ) GO TO 422                                       CON02090
      JJ = JJ + NB                                                      CON02100
      GO TO 423                                                         CON02110
  422 CALL MOVBYT(IOBUF, IBASE, 1, IOBUF, JJ, 1, NB)                    CON02120
      JJ = JJ + NB                                                      CON02130
  423 IF(IKNT .EQ.NCHAN) GO TO 421                                      CON02140
  416 CONTINUE                                                          CON02150
C                                                                       CON02160
C *****  INTERPOLATE INTO THE CYLINDER                                  CON02170
C                                                                       CON02180
  421 DO 500 ICHN = 1, NCHAN                                            CON02190
      DO 430 ICOL = NFCOL, NLCOL                                        CON02200
      JJ = ICBASE + (ICOL - NFCOL) * (NNCOL + 1) + 1                    CON02210
      IBC = ICOL - NFCOL - 1 + INBASE                                   CON02220
      X = 0.0                                                           CON02230
      DO 417 J = 1, INPCOL                                              CON02240
      I2 = 0                                                            CON02250
      NZ(2) = ICBUF(IBC + J)                                            CON02260
      X = X + LCEFF(J) * I2                                             CON02270
  417 CONTINUE                                                          CON02280
      II = X                                                            CON02290
      CYLIN(JJ) = NI(2)                                                 CON02300
      DO 429 KK = 1, NNCOL                                              CON02310
      KJ = KK + INPCOL                                                  CON02320
      X = 0.0                                                           CON02330
      DO 418 J = 1, INPCOL                                              CON02340
```

FILE. . . CONVO FORTRAN P2

```
      I2 = 0                                                            CON02360
      NZ(2) = ICBUF(IBC +J)                                             CON02370
      X = X + LCEFF(KKJ + J) * I2                                       CON02380
  418 CONTINUE                                                          CON02390
      II = X                                                            CON02400
      CYLIN(KK + JJ) = NI(2)                                            CON02410
  429 CONTINUE                                                          CON02420
  430 CONTINUE                                                          CON02430
      INBASE = INBASE + NSAMP                                           CON02440
      KL = ((IBOT - 1) * NCHAN + ICHN - 1) * 6                          CON02450
      DO 440 I = 1, 6                                                   CON02460
      CALIBC + I) = IOBUF(INBASE + I)                                   CON02470
  440 CONTINUE                                                          CON02480
      INBASE = INBASE + 6                                               CON02490
      ICBASE = ICBASE + ICYLLN                                          CON02500
  500 CONTINUE                                                          CON02510
      GO TO 211                                                         CON02520
      END                                                               CON02530
```

2.7 Forestry Applications of Computer Aided Analysis Techniques

INTRODUCTION

Background

The involvement of LARS staff in the Forestry Applications Project has been directed toward development and documentation of computer-aided analysis procedures for forest mapping. The principal activity during this year has been in documenting an approach to the selection of training areas. Advances have also been made in defining a statistical approach to the evaluation of classification accuracy. Additional activity occurred in the areas fo wavelength band selection, multitemporal analysis and change detection. Lastly, an opportunity arose to modify the computer classification output so that it would closely resemble a map product. As in the past, the primary test site for this activity has remained the Sam Houston National Forest. Results will also be reported for data collected over the San Juan National Forest in Colorado and the Hoosier National Forest in Indiana.

The material presented in this report summarizes the significant findings to date. However, there are still areas of study where more work is required. These are identified in Table 2.7-1, the Research Task Matrix.

Table 2.7-1 RESEARCH TASK MATRIX

TASKS	Current Research Status				
	Objectives	Approach	R & D	T & E	Documentation
1. Training Area Selection "Modified Cluster	X	X	X	X	X
To develop and document procedures for selecting from MSS data training areas for performing forest cover classifications					
2. Statistical Evaluation	X	X	X	X	
To develop and document a suggested set of procedures to follow for evaluating computer classifications of forest sites.					
3. Wavelength Band Selection	X	X	X	X	X
To select and evaluate which wavelength bands or regions are most suitable for classfying forest cover with MSS data.					
4. Multi-Date Analysis Procedures	X	X	X		
Develop a set of procedures which are best adapted for using multitemporal data for classification of MSS data over forested test sites.					
5. Change Detection	X	X	X		
Investigate the various approaches to performing change detection with LANDSAT data.					
6. Line maps	X	X	X	X	
Develop techniques to produce line rather than grayscale maps.					

OBJECTIVE

The overall objective of this research task was: "To work toward refining and documenting procedures on how to use various computer-aided analysis techniques in an operational forestry mapping application.

MATERIALS AND METHODS

LARSYS

Cognizant of FAP's desire to transfer new technology to the U.S. Forest Service, LARS staff working on this project refrained from developing new computer programs whenever possible. However, new capabilities often require new programming. The capability to utilize a CalComp Plotter to produce line maps is an example of an area where new programming was required. Programming activities were kept to a minimum and standard programs which appear in the LARSYS 3.1 DOCUMENTATION or are available on the LARSYS Experimental Library were utilized.

Data Availability

The following data sets were utilized for analysis:

1. LANDSAT-1 data, multiple dates, for both the Sam Houston National Forest and the Hoosier National Forest.

2. NC-130 aircraft data, MSDS 24 channel scanner, collected over part of the San Juan National Forest during NASA Aircraft Mission 247.

Test Sites

1. Sam Houston National Forest is located approximately 50 miles north of Houston on the East Texas Coastal Plain. Predominantly a southern pine mono type.

2. Hoosier National Forest is located in south central Indiana on undulating Illinoian aged flacial deposits. The oak-hickory cover predominates to the extent that the site is almost a hardwood mono type.

3. The San Juan National Forest is located in south west Colorado on uplifted volcanic material that has been extensively altered by glaciation. Conifer, both pines and spruces and aspen are the predominant cover type.

PROCEDURES

The material presented in this section deals with the task statements defined within the overall research objective. These tasks can be briefly stated as:

1. Selection of training sets.
2. Statistical evaluation of classification accuracy.
3. Wavelength band selection.
4. Wavelength band selection for multitemporal data sets.
5. Change detection

Selection of Training Sets

The publication "A Forestry Application Simulation of Man-Machine Techniques for Analyzing Remotely Sensed Data",[1] was prepared specifically for this task. Personnel from the FAP and Technology Transfer activity worked together to produce this training simulation. In addition to addressing the question of training set selection for computer-aided analysis, the authors go into more detail regarding the entire concept of machine-assisted analysis.

This document is aimed at an audience that has not been extensively involved in remote sensing activities. The concepts described are aimed at giving the reader a new perspective. This publication could be considered the first in a series of successive volumes each more detailed. Such detail could include in-depth studies of certain analysis procedures or specific applications examples, such as range, water or timber management.

Evaluation of Classification Accuracy

Discussion

The objective of this task was to develop procedures to statistically evaluate the accuracy of computer-assisted LANDSAT classifications. Results reported in "Analysis of Aircraft MSS Data for Timber Evaluation" (Mroczynski, et.al. 1976), and in the FAP final report for CY75 indicate the suitability of statistical evaluation of classification accuracy for large geographic areas. A necessary requirement for performing this evaluation is that the analyst have good ground reference data available. Without current ground reference data, either aerial photographs and/or detailed ground cover maps, the evaluation cannot be as meaningful.

Unfortunately, our experience has been that accurate ground information is not always available for forested test sites. Forest type maps generally indicate the age, size and merchantable status of a forest stand.

These qualities may not relate to the stands spectral characteristics which are the basis for the statistical evaluation, so therefore may or may not be meaningful ground reference information.

Providing suitable ground reference information is available, a systematic sample grid can be aligned with the classification. The analyst then interprets these fields with the aid of the ground reference information. When the fields have been interpreted, various statistical tests can be applied to the results and inferences can be drawn about the classification accuracy for the entire site. This section will emphasize the statistical test procedures.

Figure 2.7-1 outlines the steps in the statistical evaluation procedure. An underlying assumption in this figure is that there exists a well-defined analysis objective. After obtaining the preliminary classification, the analysis should question how well it meets the analysis objectives. If the analysis is confident in his results there may be no need for further evaluation. Undoubtedly, the data analyst is not the ultimate user of the classification information. The statistical evaluation may be used to capture the user confidence and therefore may be performed even if the analyst was confident in his results.

The essential steps in the evaluation are:

- List the cover types to be tested
- locate potential test fields by systematic sampling with random start
- interpret and label the acceptable test fields
- perform the analysis of variance (including investigation of interaction for 2-factor Analysis of Variance) and range test calculations.

The tables which follow illustrate the Analysis of Variance procedures and are set up in the general form of 1) verbal description of step, and 2) numerical example. These tables are applicable in any situation by substituting values and performing the new evaluations based on the new values.

Analysis of Variance Tests

Remote sensing MSS data are fundamentally binomial in nature. Pixels in computer classifications are either identified correctly or incorrectly, hence their binomial distribution. As a result, the arcsin $\sqrt{p}$ transformation should be applied since, according to Steele and Torrie (1960):

> "The data can be transformed or measured on a new scale of measurement so that the transformed data are approximately normally distributed. Such transformations are also intended

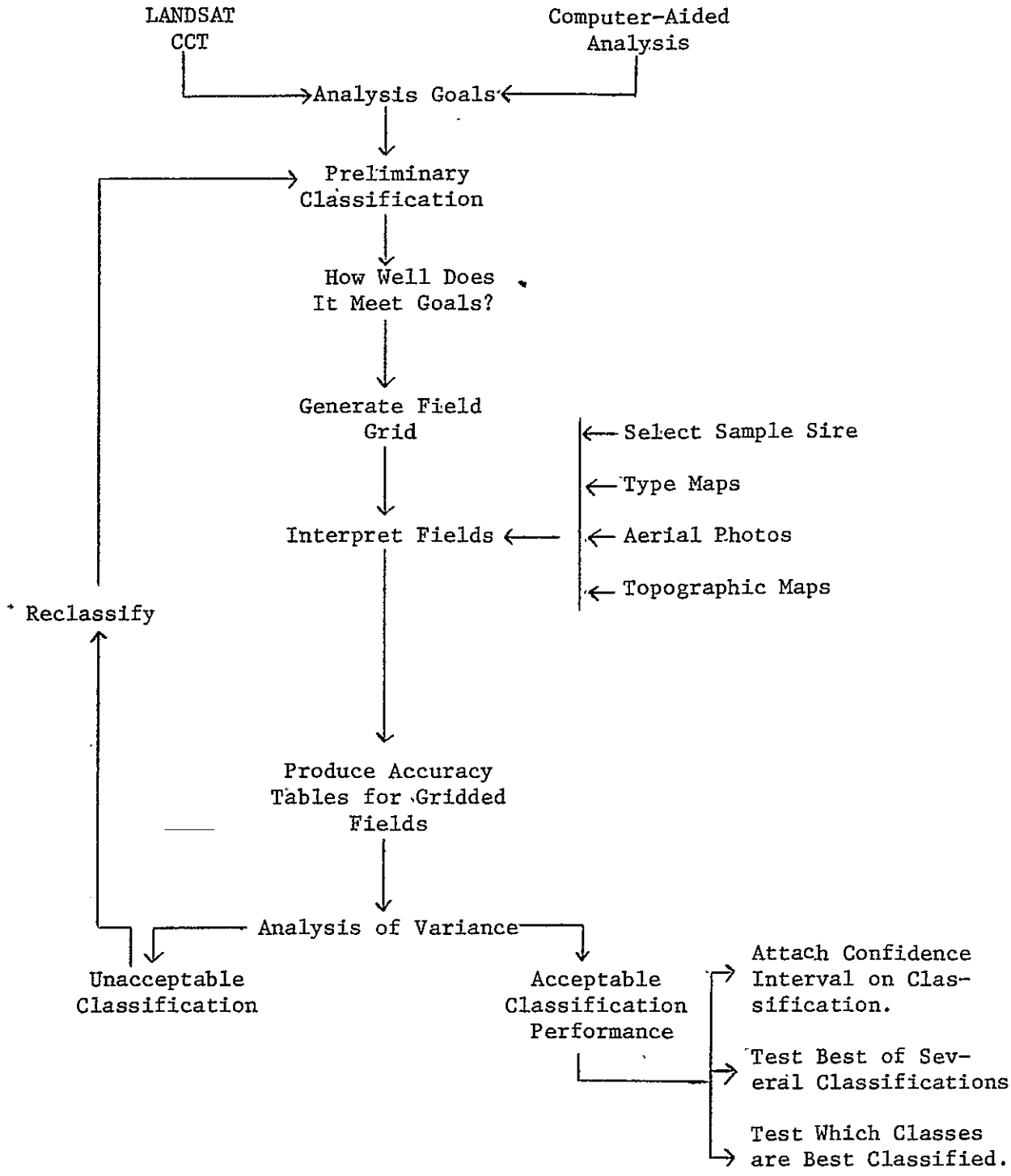


Figure 2.7-1 Steps in Implementing the Stage II Statistical Evaluation Technique

to make the means and variances independent, with the resulting variances homogeneous. This result is not always attained."

The importance of making the means and variances independent and the variances homogeneous is tied into the fact that these basic assumptions are made when performing an analysis of variance.

Regarding the effect of sample size upon homogeneity of variances, we recommend that at least 50 to 100 observations (one pixel = one observation) be obtained for each cover type to be tested. This removes the need for application of corrections to small sample sizes (50) as recommended by Snedecor and Cochran (1967). For sample sizes ranging from 50 to 1000, the comparisons among the percentage accuracy of cover type identification may be somewhat influenced by unequal variances. But for most studies an adjustment or weighting by the actual sample size is very seldom needed to obtain reasonably good comparisons. Especially if the range is from 100 to 500 samples, the assumption of homogeneity of variances is not usually violated enough to warrant a weighted transformation before running an ANOVA.

The main advantages of the angular ($\arcsin\sqrt{p}$) transformation are that the error variance of the resulting observations (in degrees) is approximately constant, has infinity degrees of freedom (∞ df), and is equal to 821/n ("n"; sample size). The transformation is used as the unbiased estimator of the mean square error. Since sample sizes will vary among cover types, the harmonic mean (Table 2.7-7, which averages the different numbers of observations per accuracy mean, should be used according to Steele and Torrie[3].

Two analyses of variance which will be encountered most often in LANDSAT applications are: 1) one-factor ANOVA, and 2) two-factor ANOVA. The one-factor ANOVA would be used to test for significant differences among cover types of a single classification or among overall accuracies of different classifications of the same data (e.g., wavelength band studies). The model for the one-way ANOVA (assuming transformed accuracy means) is:

$$Y_i = \mu + C_i + E(i)$$

where

Y_i = the overall accuracy (in degrees) of the i-th classification or cover type

μ = true overall accuracy mean

C_i = effect of the i-th classification or cover type

$E(i)$ = random error of the i-th classification or cover type

The best estimator of $E(i)$ is assumed to be 821/n (n= harmonic mean of sample sizes). This estimator is used as the mean square error

(denominator, with infinity degrees of freedom) in the F test for significant variation (Tables 2.7-2 and 2.7-3.)

A two-factor analysis of variance would entail testing for significant differences among different classifications of the same data set (e.g., wavelength band studies) and, at the same time, among cover types in the same data set. Thus, both factors (classification and cover type) are tested simultaneously. The model for a two-factor ANOVA is:

$$Y_{ij} = \mu + C_i + T_j + CT_{ij} + E(ij)$$

where

Y_{ij} = classification accuracy (in degrees) of the i-th classification for the j-th cover type

μ = true overall accuracy mean

C_i = effect of the i-th classification

T_j = effect of the j-th cover type

CT_{ij} = effect of the interaction between the i-th classification and the j-th cover type

$E(ij)$ = random error, which is normally and independently distributed with mean = 0 and variance - σ^2

If the interaction effect, CT_{ij}, is found to be nonsignificant, $E(ij)$ (which equals 821/n) provides the error mean square for the denominator of the F test. Again, it has infinity degrees freedom, thereby enabling a powerful F test.

Since interaction can occur between classifications and cover types, the interaction must be investigated to determine whether it is a "significant" source of variation. Essentially, an attempt is being made to find the best estimator of the error mean square for the F test. The interaction investigation proceeds as follows (and, Table 2.7-6):

A) If $(CT_{ij}/df)/821/n)$ is not significant at $\alpha = .25$ (F test), conclude there is no significant interaction and use 821/n for all F tests.

B) If it is significant at the .25 level, this may be due to the 821/n being too small because error other than binomial to normal is not included in 821/n. Hence, obtain the mean square for non-additivity with one degree of freedom from Anderson and McLean[4].

1) If the residual mean square is not significantly different at $\alpha = .25$, using 821/n as the denominator in the F, then use

821/n with infinity df for all tests.

2) If the residual mean square is significant at $\alpha = .25$, use the residual mean square for all tests. This mean square has finite degrees freedom and provides a less powerful test than 821/n with infinity df.

The Newman-Keuls Range Test is an appropriate test for discerning which accuracy means are significantly different (Table 2.7-7). For this and preceding ANOVA's, we recommend that the level of significance be set at 90% (0.1 alpha). Thus, the tests will be quite "liberal" in the sense that if any significant differences exist, they will probably be detected. Theoretically, this means that the beta (B) error is made low (error of not detecting a significant difference when it truly exists) at the expense of raising the alpha error (error of denoting significant differences when not truly present).

Test

The one-factor analysis of variance test was applied to results obtained for the Sam Houston National Forest. Table 2.7-8 shows the calculations and the Newman Keuls Range Test for five differenc cover types. The conclusions drawn from this test is that at the 90% level there is no significant difference between classes one and four, or non-forest and pine. In other words, these classes would be difficult to separate whereas the remaining classes would be more easily separable. However, the percent of the pixels correctly identified in each class is not high. One would expect better accuracy results considering the species compostion of the area.

What this example typifies is a situation where sufficient ground reference data were not available for the analyst's evaluation. Given the information that was available only a limited number of systematic fields could be evaluated as indicated by the small number of pixels in each class. If a complete set of aerial photographs were available for the entire Sam Houston Test Site, the results of the statistical evaluation might be different.

Wavelength Band Evaluation Studies

Discussion

One important consideration prior to undertaking an analysis task relates to the type of data available. The timeliness, quality, and data channels available for analysis are factors that must be considered. For this particular task we attempted to identify data channels best suited for forest mapping purposes. We were concerned with two situations:

1. A wide range of channels (MSDS 24 channel scanner of SKYLAB S-192 data) available for a single data collection date, and

Table 2.7-2 One-factor analysis of variance and Newman-Keuls Range Test, three different cover types from the same computer classification.

PROCEDURE	EXAMPLE
Apply arcsin $\sqrt{p}$* transformation to cover type classification accuracies	COVER TYPE: Agricultural (A) — NUMBER OF PIXELS: 150 — ACCURACY (%): 88.7 — TRANSFORMATION: 70.3 Forest (F) — 70 — 81.4 — 64.5 Water (W) — 80 — 98.8 — 83.6
Calculate cover type sum of squares	$SS_T = [(70.3)^2 + (64.5)^2 + (83.6)^2]/1^+ - [(70.3 + 64.5 + 83.6)^2/3^@]$ $= 191.9$
Determine cover type mean square	Mean Square (i.e., MS_T) = SS_T/ (Number of means -1) $MS_T = 191.9/2 = 96.0$
Calculate F test and determine whether significant	A) $F = 96.0/[821/89.7^{**}] = 10.5$ (significant) B) Tabular $F_{2,\infty}^{++} = 2.30$ (90% level, $\alpha = 0.1$)
Arrange transformed means in descending order	(W) 83.6 (A) 70.3 (F) 64.5
Calculate standard error of mean	$S\bar{y} = \sqrt{\text{error mean square/number obs. per mean}}$ $= \sqrt{[821/89.7]/1}$ $= 3.03$
Determine tabular ranges (Newman-Keuls Range Test)	Number-of-means range = (Studentized Range $_{(df=\infty)}$*)$(S\bar{y})$: 2 3 $R_3 = (2.902)(3.03) = 8.8$ $\alpha = 0.1$ .05=2.772 3.314 $R_2 = (2.326)(3.03) = 7.0$
Draw bars between means with ranges less than the corresponding tabular ranges	(W) 83.6 (A) 70.3 (F) 64.5 (Hence, classification accuracy of water is significantly better than that for agriculture and forest)

*These values can be found in the Appendices of most statistics texts.
+Observations per accuracy mean = 1.
@Number of accuracy means.
**Harmonic mean = number of means/Σ(1/observations per mean)= 3/(1/150 + 1/70 + 1/80) = 89.7.
++Number of accuracy means -1 = degrees freedom.

Table 2.7-3 One-factor analysis of variance and Newman-Keuls Range Test, three different classifications of same data set.

PROCEDURE	EXAMPLE
Apply arcsin $\sqrt{p}$ transformation to overall classification accuracies	CLASSIFICATION: 1 — OVERALL ACCURACY (%): 88.3 — TRANSFORMATION (DEGREES): 70.0 2 — 80.7 — 63.9 3 — 89.7 — 76.3
Calculate classification sum of square	SS_c = [(70.0)^2 + (63.9)^2 + (71.3)^2]/1* − [(70.0 + 63.9 + 71.3)^2/3+] = 31.2
Determine classification mean square	Mean Square (i.e., MS_c) = Sum square (i.e., SS_c)/(Number of means −1) MS_c = 31.2/2 = 15.6
Calculate F test and determine whether significant	A) F = 15.6/ [821/300@] = 5.7 (significant) B) TABULAR $F_{2,\infty}$++ = 2.30 (90%)
Arrange transformed means in descending order	(3) (1) (2) 71.3 70.0 63.9
Calculate standard error of mean	$S_{\bar{y}}$ = $\sqrt{\text{error mean square/number observations per mean}}$ = $\sqrt{[821/300\ @]\ /1}$ = 1.65
Determine tabular ranges	Number-of-means range = (Studentized range)$_{(df=\infty)}$ ($S_{\bar{y}}$) (90%) R_3 = (2.902)(1.65) = 4.8 R_2 = (2.326)(1.65)= 3.8
Draw bars between means with ranges less than the corresponding tabular ranges	(3) (1) (2) 71.3 70.0 63.9 (Hence, the second classification is significantly different from the first and third.)

*Observations per accuracy mean = 1.
+Number of accuracy means.
@There are 300 test pixels (observations).
++Number of accuracy means −1 = degrees freedom.

Table 2.7-4 Arcsin$\sqrt{\hat{p}}$ transformation of classification results. This transformation changes the binomial nature of these values (pixels are correctly or incorrectly identified) to a new scale of measurement so that the assumptions necessary for analysis of variance can be made. The transformed data should be approximately normally distributed, means and variances independent, and the resulting variances homogeneous (4).

CLASSIFICATION	COVER TYPE	NUMBER OF OBSERVATIONS	PERCENTAGE CORRECT CLASSIFICATION, $\hat{p}$	ARCSIN $\sqrt{\hat{p}}$ (degrees)
1	Agriculture (A)	150	86.7	68.6
	Forest (F)	70	85.7	67.8
	Water (W)	80	93.7	75.5
2	Agriculture	150	80.0	63.4
	Forest	70	74.2	59.5
	Water	80	87.5	69.3
3	Agriculture	150	88.6	70.3
	Forest	70	81.5	64.5
	Water	80	98.8	83.6

Table 2.7-5 Calculation of sums of squares.

CALCULATION CATEGORY	CALCULATION
Correction term, C.T.	$CT = (68.6 + 67.8 + \ldots + 83.6)^2/9^*$ $= 43,056.3$
Classification	$SS_c = [(68.6 + 67.8 + 75.5)^2 + (63.4 + 59.5 + 69.3)^2 + (70.3 + 64.5 + 83.6)^2]/3^+ - C.T.$ $= 124.0$
Cover type	$SS_T = [(68.6 + 63.4 + 70.3)^2 + (67.8 + 59.5 + 64.5)^2 + (75.5 + 69.3 + 83.6)^2]/3^@ - C.T.$ $= 236.7$
Total	$SS_{TOT} = [(68.6)^2 + (67.8)^2 + \ldots + (83.6)^2] - C.T.$ $= 43,456.9 - 43,056.3$ $= 400.6$
Interaction and/or error	$SS_{INT/ER} = SS_{TOT} - (SS_T + SS_c)$ $= 400.6 - (236.7 + 124.0)$ $= 39.9$
Set up ANOVA table	(see below)

source of variation	degrees of freedom (df)	sum of squares (SS)	mean square (SS/df)
Classification	(Classifications -1) = 2	124.0	62.0
Cover type	(Cover types -1) = 2	236.7	118.4
Interation and/or error	8-(2+2) = 4	39.9	10.0
TOTAL	Accuracy means-1 = 8	400.6	

* Number of means (cells).
\+ Number of cover types.
@ Number of classifications.

Table 2.7-6 Investigation of Interaction.

PROCEDURE	EXAMPLE
A) Test for significance of interaction- STOP HERE if not significant-use 821/n for all tests	(75% level, $\alpha = .25$) F = Interaction and/or error mean square/$[821/n^*]$ = 39.9/[821/89.7] = 4.3 (significant) tabular $F_{4,\infty}^{+} = 1.35$
B) Obtain the mean square for Non-additivity with 1 degree of freedom	1) Produce following table:

CATEGORY	COVER TYPES AG	FOR	WAT	ROW MEANS	ROW MEAN- OVERALL MEAN	(ROW MEAN- OVERALL MEAN)^2
Cl. 1	68.6	67.8	75.5	70.63	1.46	2.13
Cl. 2	63.4	59.5	69.3	64.07	-5.10	26.01
Cl. 3	70.3	64.5	83.6	72.80	3.63	13.18
Col. $\bar{x}$	67.43	63.93	76.13	$\bar{\bar{x}} = 69.17$		$\Sigma = 41.32$
Col. $\bar{x}$- overall $\bar{x}$	-1.74	-5.24	6.96			
(")^2	3.03	27.46	48.44	$\Sigma = 78.93$		

2) Calculate mean square (i.e., sum of squares/1)= $[(68.6)(1.46)(-1.74)+(67.8)(1.46)(-5.24)+(75.5)(1.46)(6.96)+(63.4)(-5.1)(-1.74)+(59.5)(-5.1)(-5.24)+(69.3)(-5.1)(6.96)+(70.3)(3.63)(-1.74)+(64.5)(3.63)(-5.24)+(83.6)(3.63)(6.96)]^2/(41.32)(78.93) = 13.81$

B1) Test residual mean square for significance- STOP HERE if not significant- use 821/n for all tests	1) MS_{res} = Interaction and/or Error SS - Non-additivity MS/ residual degrees freedom = [39.9- 13.81] /3@ = 8.7 2) F = 8.7/$[821/n^*]$ (75% level) = 8.7/9.2 = 0.94 (not significant) tabular $F_{3@,\infty} = 1.37$
B2) Use residual mean square for all tests	MS_{res} = 8.7, df = 3@

*Harmonic mean = number of accuracy means/Σ(1/observations per mean) = 9/(1/150 + 1/150 + . . . + 1/80) = 89.7.

+Degrees of freedom for interaction and/or error mean square.

@Degrees of freedom for residual sum of squares = degrees freedom for interaction and/or error SS - 1.

Table 2.7-7 Tests for significance using best estimator of error mean square, as determined by investigation of interaction.

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F TEST	TABULAR F (90%)
Classification	(Classifications- 1)= 2	124.0	62.0	(signif.) 6.7	$F_{2,\infty}$= 2.30
Cover Type	(Cover types- 1)= 2	236.7	118.4	(signif.) 12.9	$F_{2,\infty}$= 2.30
Error*	∞	———	827/89.7+ = 9.2		

*The residual mean square was found to be nonsignificant. If significant, the residual mean square would be used in place of 821/n for all tests.

+Harmonic mean.

Table 2.7-8 Newman-Keuls Range Test for classifications and cover types. Even if only one of two main factors is significant, the means may be kept separate for the range tests.

PROCEDURE	EXAMPLE
Arrange transformed means in descending order	(3-W)* (1-W) (3-A) (2-W) (1-A) (1-F) (3-F) (2-A) (2-F) 83.6 75.5 70.3 69.3 68.6 67.8 64.5 63.4 59.5
Calculate standard error of mean	$S_{\bar{y}} = \sqrt{\text{error mean square/observations per cell}}$ $= \sqrt{9.2 / 1} = 3.03$
Determine tabular ranges	Number-of-means range = (Studentized range)$_{(df+=\infty)}$ $(S_{\bar{y}})$ (90% level, 0.1) $R_9 = (4.037)(3.03) = 12.2$ $R_8 = (3.931)(3.03) = 11.9$ $R_7 = (3.808)(3.03) = 11.5$ $R_6 = (3.661)(3.03) = 11.1$ $R_5 = (3.478)(3.03) = 10.5$ $R_4 = (3.240)(3.03) = 9.8$ $R_3 = (2.902)(3.03) = 8.8$ $R_2 = (2.326)(3.03) = 7.0$
Draw bars between means with ranges less than appropriate tabular ranges.	83.6 75.5 70.3 69.3 68.6 67.8 64.5 63.4 59.5

*Accuracy mean for water of third classification.

+Same degrees freedom as the error (residual) MS. If the residual MS in Table 5 had been significant, the df for Studentized range would be 3.

Table 2.7-9

PROCEDURE	EXAMPLE
apply $\arcsin\sqrt{p}$	COVER TYPE / NUMBER OF PIXELS / ACCURACY (%) / DEGREES 1 non-forest 144 69.4 56.42 2 hardwood 56 82.1 64.97 3 mix 72 34.7 36.09 4 pine 348 59.2 50.30 5 dense pine 180 22.2 28.11
SS	$SS = [(56.42)^2 + (64.97)^2 + (36.09)^2 + (50.30)^2 + (28.11)^2]/1 - [(56.42 + \ldots + 28.11)/5]$ $= [3.83.2 + 4221.1 + 1302.5 + 2530.1 + 790.2]/1 - 11,128.8$ $= 12,027.1 - 11,128.8 = 898.3$
MS	$898.3/(5-1) = 224.6$
F Test	$F = 224.6/[821/(5/(1/144 + 1/56 + 1/72 + 1/348 + 1/180))]$ $= 224.6/[821/106.0]$ $= 29.0$ (significant) tabular $F_{4,00} = 1.94$ (90% level, $\alpha = 0.1$)
$s_{\bar{y}}$	$s_{\bar{y}} = \sqrt{7.75/1}$ $= 2.78$
tabular ranges (N-K)	$R_5 = (3.478)(2.78) = 9.67$ $R_4 = (3.240)(2.78) = 9.01$ $R_3 = (2.902)(2.78) = 8.07$ $R_2 = (2.326)(2.78) = 6.47$
	(2) (1) (4) (3) (5) 64.97 56.42 50.30 36.09 28.11 (0.1 90%)
	82.1% 69.4% 59.2% 34.7% 22.2% hardwood non-forest pine mix dense pine

2. A narrow range of channel (LANDSAT) but available for multiple points in time.

There are unique problems associated with each case. In the first the analyst is concerned with selecting a meaningful channel set that will yield a good classification without using an excessive amount of computer time. The time factor is critical in areas where large amounts of data must be processed. Generally fewer channels require less processing time.

The second situation requires selecting the best channel of combination of channels from data sets collected at different times. The interactions that must be considered are both occurring within and between data sets. This situation generally occurs with satellite collected data which are easily overlayed in registration, as in the case of LANDSAT data.

Single Data Analysis

The single data analysis is probably the most common data classification situation attempted. If LANDSAT data is involved, there is not much concern about selection of channels, since only four are available. However, with aircraft and future satellite systems, the concern for selecting the "best" channel from a larger number of channels is understandable.

For this study MSDS 24-channel scanner data collected over the San Juan National Forest as part of NASA Mission 247 was utilized. Figure 2.7-1 indicates the general location of the study site and the composition of the ground cover. These data were collected August 4, 1973.

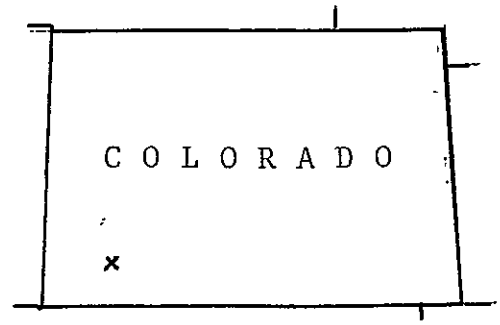


Figure 2.7-1 Approximate test site location in Southwestern Colorado. Cover type composition and caption is shown in Figure 2.7-2.

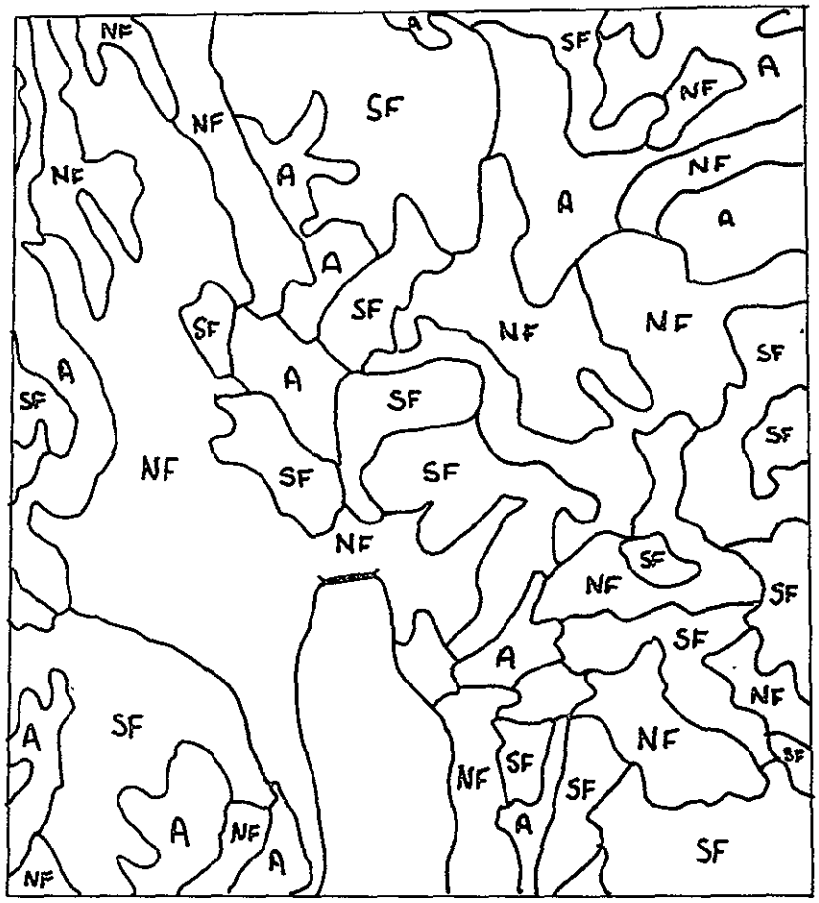


Figure 2.7-2 Generalized forest type map of the test site. The classes have been simplified from a Forest Service map to indicate the species composition of the site. The cover types are:

A - Aspen
SF - Spruce-Fir
NF - Non-forest, including non-commercial tree species, shrub la bare soil and rock

The map includes a nine section area around the Rio Grande Reservoir.

Our objective was to select the channel combinations that would yield the best classification for general forest cover types. Realizing the future significance of the Thematic Mapper (TM) we limited our investigation to a subset of the 24 channels that closely resembled the TM channels. Table 2.7-10 indicates the proposed TM channels and the MSDS channels. The boxes aroung the MSDS channels indicate which were used for this study.

Training sets were picked within the study area that we felt were representative of the cover types. These estimations were based on coincident aerial photography collected during Mission 247 and Forest Service cover type maps. The scanner data were processed with the *SEPARABILITY function of LARSYS. Table 2.7-11 gives the channel combinations for the best 2 through 7 channels. Similar results over the same test site were obtained from SKYLAB S-191 data and reported by Hoffer et al: [5].

All seven channels were used to classify a small poriton of the study area surrounding the Rio Gradne Reservoir. Figure 2.7-3 shows a CalComp map of the results of that classification and a black-and-white copy of a color infrared aerial photo from Mission 247. This classification can also be compared to Figure 2.7-2 which is the forest service type map of the same area. The classification has good visual agreement when compared with the aerial photo. A certain amount of disagreement can be expected between the classification and type map because of the different criterion used to develop each map. The classification is based on spectral criteria while the type map is based on timber management requirements.

Once a satisfactory classification for the ten classes of material had been developed successive classifications, each deleating channel and using the next "best" selection of channels were produced. This procedure was followed to determine the point at which the classification accuracy for forest falls below the point of being acceptable. Since we had no absolute measure of classification accuracy and we were well satisfied with the visual correspondence of the seven channel classification and the air photo, we assumed this classification to be the best.

For an actual mapping situation, the analyst would attempt to minimize the number of channels used to save processing costs. Previous studies at LARS (Coggeshall and Hoffer, 1973[6]) indicate that past five channels the increase in classification accuracy is not enough to warrant the increased processing cost.

The area estimates for each class of the five classifications were compared against the estimated area of those cover types as mapped on the seven channel classification. Table 2.7-12 gives the percent area for each cover type for the various classifications. A review of the table indicates that there were no drastic changes as the numbers of channels were decreased. If this relationship is valid for other data sets in other geographic areas it would provide for useful guidance for appling

Table 2.7-10 Proposed Thematic Mapper Channels

Bendix 24-channel Data From Mx 247 San Juan Test Site		Proposed Thematic Mapper Channels LANDSAT-D	
ch. no.	Bandwidth	ch. no.	Bandwidth
24	.38-.40		
12	.40-.44		
1	.47-.49	1	.45-.52
13	.53-.58	2	.52-.60
2	.59-.64		
14	.65-.69	3	.63-.69
3	.72-.76	4	.74-.80
15	.77-.81		
4	.82-.88	5	.80-.91
16	.98-1.04		
23	1.06-1.09		
11	1.13-1.17		
5	1.20-1.30		
17	1.53-1.62	6	1.55-1.75
6	2.30-2.43		
18	3.78-4.04		
7	4.05-4.46		
19	6.00-7.00		
8	8.27-8.70		
20	8.80-9.30		
9	9.38-9.88		
21	10.10-11.0	7	10.4-12.5
10	11.00-12.0		
22	12.00-13.4		

Table 2.7-11

CHANNEL SELECTION *

GRANDE RESERVOIR TEST SITE

Channels and Wavelengths	Channel Combinations					
	2 channels	3 channels	4 channels	5 channels	6 channels	7 channels
Channel 1 0.47-0.49		x	x	x	x	x
Channel 3 0.72-0.76				x	x	x
Channel 4 0.82-0.88		x		x	x	x
Channel 13 0.82-0.88					x	x
Channel 14 0.65-0.69			x	x	x	x
Channel 15 0.77-0.81	x	x	x			x
Channel 21 10.1-11.0	x		x	x	x	x

* Results obtained from utilization of LARSYS (Version 3.2) function *SEPARABILITY The channels listed are a selected subset of 24 channels collected during NASA Mission 247 flown August 4, 1973.

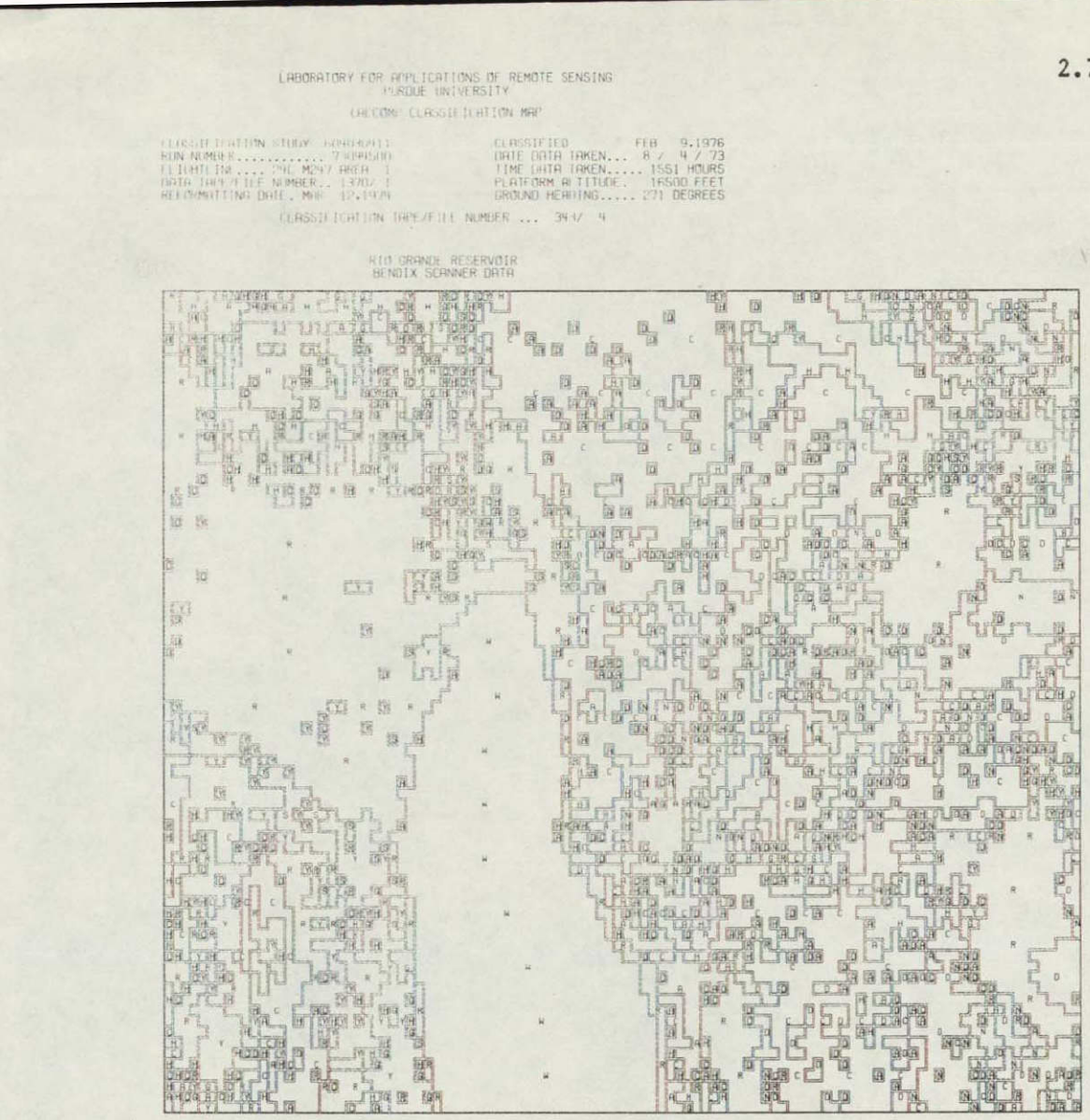


Figure 2.7-3 CalComp classification of the test site corresponding to the area in Figures 2.7-3 and 2.7-5. The classes are:

- C - Conifer, spruce-fir
- A - Aspen
- S - Bare soil
- Y - Grass and soil mixture
- G - Grass
- H - Shrub areas
- R - Bare rock
- W - Water
- D - Dense conifer
- N - Topographic shadow

Because this map has been generated from classification of aircraft data the scale of the map varies across the map.

Aerial photo in Figure 2.7-4 corresponds to area in map.

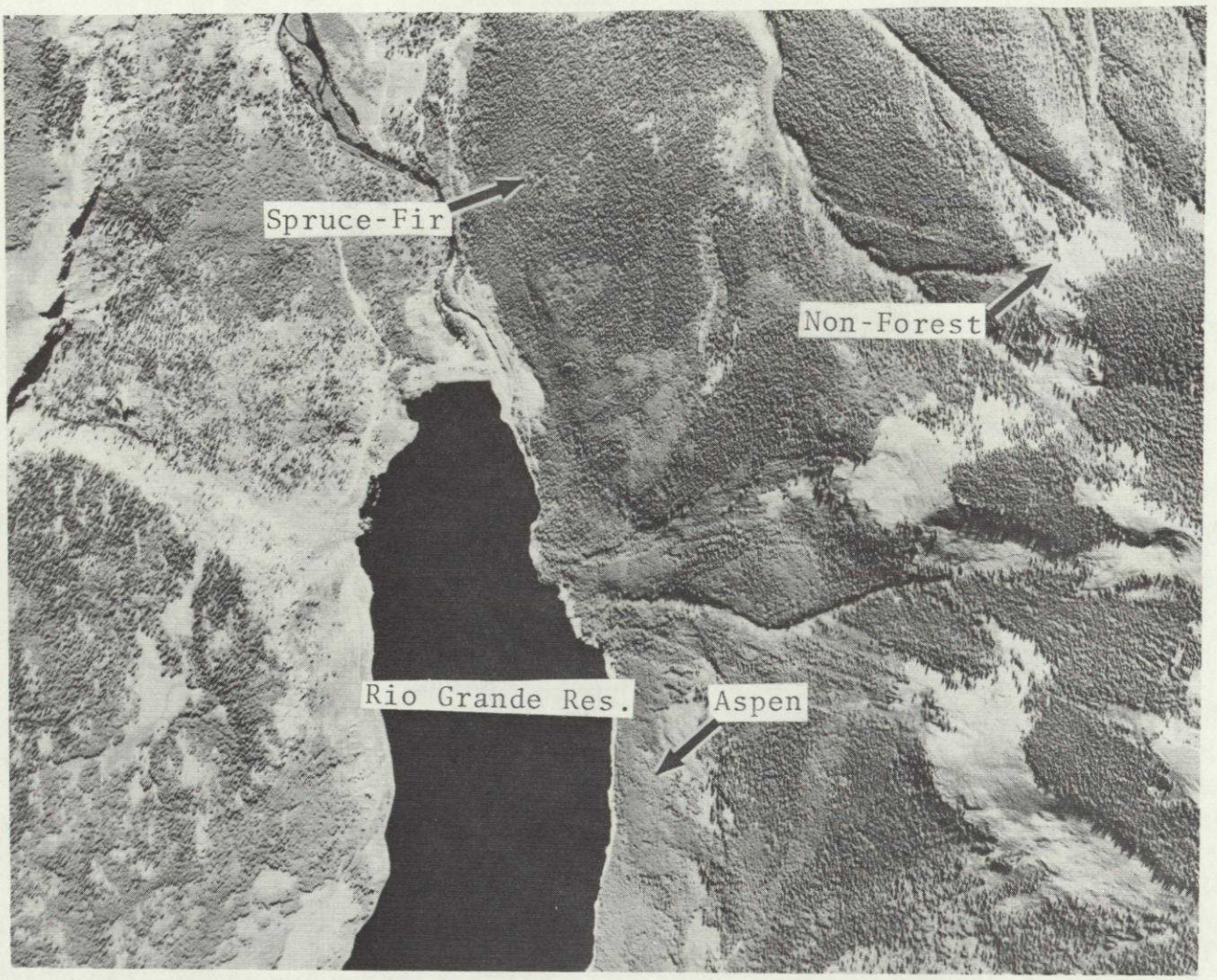


Figure 2.7-4 Black-and-white copy of a color infrared transparency from NASA Mission 247. Representative classes of Aspen (A), Spruce-Fir (SF) and Non-Forest (NF) have been marked which correspond to similar areas on the classification.

Table 2.7-12 Channel Selection Study

COVER TYPES

NUMBER OF CHANNELS USED IN CLASSIFICATION	DENSE CONIFEROUS TREES	NON DENSE CONIFEROUS TREES	ASPEN	CHAPARREL	GRASS	GRASSY SOIL	BARE SOIL	ROCK	WATER	SHADOW
7 CHANNELS	11.1	26.4	7.9	8.8	1.3	4.6	1.5	24.2	11.5	2.7
6	11.8	25.4	8.2	8.9	1.4	4.6	1.5	24.1	11.5	2.7
5	12.1	26.1	7.3	9.0	1.5	4.5	1.4	24.0	11.6	2.6
4	13.7	24.9	8.1	8.9	1.3	4.5	1.1	23.5	11.5	2.5
3	14.1	23.8	8.7	9.0	1.4	4.7	0.9	23.3	11.6	2.6
2	13.1	27.6	5.1	8.4	1.4	4.3	3.5	22.2	11.6	2.8

CHART SUMMARIZING CLASSIFICATION RESULTS

RIO GRANDE RESERVOIR SUBSET

*Numbers are in percent of the total data points of the area classified.

satellite data to attempting parge area analysis. This would be especially true if the analysis objects required specific information on general forest cover, an area not drastically affected by a reduction in the number of channels.

Additional work along with simulated Thematic Mapper scanner data should be pursued, especially in different geographic locations with differing mixes of forest cover types.

Multi-date Analysis

The LANDSAT satellites provide the opportunity to acquire, register and analyze multi-date data sets. This capability allows for analysis of data in which the identification of certain cover type categories is maximized (e.g., such as separating mixed stands of conifers and hardwoods). Once a multi-date data set has been created the analyst is faced with the problem of selecting channels from the data available for analysis.

The channel set selection process would seem identical to the single date situation. Indeed, the same *SEPARABILITY processor is used in either situation. The problem with multi-date data rests with the selection of the appropriate data sets to use for analysis, then further selection of channels within those data.

We had a seven date LANDSAT data set available over a portion of the Hoosier National Forest in central Indiana. Our approach was to first visually rank the data over the test site. We felt that if the data were ranked we could eliminate certain dates based on the appearance of the data at that time of year. Our ranking was based on color infrared composites produced on the digital display. Ten categories were ranked on a scale of 0-3, based on how distinguishable that class was at a given date.

Table 2.7-13 gives the visual rating for the ten classes. The higher the number, the more distinctive the class. Based on this ranking the May data set should have been the best for all cover types while the June data set the worst.

Our next step was to use the *SEPARABILITY processor to see if these results would be comparable. *SEPARABILITY was run independently for each of the seven dates to determine the channel ranking within the data sets. The best two channels for each date were then run through *SEPARABILITY to determine the best channels between data sets. This procedure was repeated for the next best two channels so that eventually all 28 channels were run through *SEPARABILITY. The data were processed as two groups of 14 rather than one group of 28 for efficiency's sake. There are many more combinations of 28 than of 14, therefore, consuming more time to produce the required information. The results of these efforts are shown in Table 2.7-14. The channel ranking appears next to the bandwidth which appears in parenthesis. The last column in the table gives the results of

Table 2.7-13 Visual Distinctiveness Matrix. Each cover type/date combination was assigned a number from low (0) to high (3) distinctiveness depending upon how easily the cover type could be identified in simulated color-infrared imagery from the LARS digital display. Three foresters assigned these numbers by group concensus.

LOCATION	DATE	COVER TYPE											SUMMATION			LANDSAT SCENE ID
		Conifer	Hardwood 1 (upland)	Hardwood 2	Hardwood 3 (lowland)	Cropland	Pasture	Water 1 (clear)	Water 2	Water 3 (turbid)	Cultural	Mean for Water	Water not average	Rank	Water averaged	
HNF-Central Indiana	May 4, 73	1	2	3	3	2	2	1	3	2	2	2	21	1	17	1285-16001
	June 8, 73	1	2	2	0	0	0	1	2	0	0	1	8	5	6	1320-15541
	Aug. 19, 73	0	1	1	0	3	3	3	2	1	2	2	16	4	12	1392-15531
	Sept. 7, 73	0	3	1	0	2	3	3	3	3	1	3	19	2	13	1411-15584
	Nov. 17, 73	3	0	0	0	3	3	3	3	3	1	3	19	2	13	1482-15514
	Feb. 15, 74	3	1	0	1	1	1	3	3	3	1	3	17	3	11	1572-15493
	Mar. 8, 74	3	1	0	1	3	2	3	3	3	0	3	19	2	13	1591-15550

Table 2.7-14 RANKING OF MULTITEMPORAL DATA SET

Date	Rank	
	Best 2 of 4 channels per date	Other 2 of 4 channels per date
	1st 14	2nd 14
May 4, 1973	7 (VIS 2) 11 (IR 2)	7 10
June 8, 1973	1 (VIS 2) 2 (IR 1)	2 1
August 19, 1973	3 (VIS 2) 13 (IR 2)	3 11
September 7, 1973	10 (VIS 2) 5 (IR 1)	6 5
November 17, 1973	12 (VIS 2) 14 (IR 1)	13 12
February 15, 1974	4 (VIS 2) 9 (IR 2)	14 4
March 6, 1974	8 (VIS 2) 6 (IR 2)	8 9

the second set of 14 channels.

A comparison of the visual and numerical ranking shows that the order of the ranking has been changed considerably. June, the lowest ranking in the visual is the best in the numerical ranking. And, in fact, subsequent analyses has indicated that June is probably a better data set for overall analysis than May.

The May data was collected early in the month before the hardwoods were completely leaked out. The variation in leaf cover combined with the varied response of the understory made the May data difficult to analyze. The understory, especially the dogwood's (Cornus florida) above the ridge top - which leaf early - inflated the visual rating, causing May to be ranked first.

On the June data the hardwoods are in their peak growth. Spectral variability is therefore minimized, and the separability of the hardwoods from the other cover types is maximized. June, however, was an excessibly wet month. Many of the agricultural and pasture land were saturated with rain water. This fact altered the colors in the composit therefore affecting the visual analysis.

The purpose of this comparison is to point out some items worthy of consideration. Namely, that once the decision has been made to proceed with a multitemporal analysis, it is not an easy matter to select the appropriate data. Creating a three or more date overlay is neither expedient or cost-effective. Visual interpretation of the data is subject to human bias and error. Additionally, the study site might not warrant a multitemporal analysis.

In the case of the Hoosier data classification accuracy was acquired for seven supervised training classes. The classes included deciduous and coniferous forest, cropland, pasture, residential, and clear and turbid water. These results are presented in Table 2.7-15. Accuracy remains relatively constant from 14 to 2 channels. The 95.5% accuracy for the best two channels represent the June data set. This would tend to indicate that for the Hoosier test site the multi temporal analysis was not necessary since a single date data set yields quite reasonable results.

Change Detection

The ability to detect and identify changes in the forest scene are important information to resource managers. The Regional Applications Project (RAP) is working on developing change detection procedures for the Texas Coastal Zone. The FAP staff at LARS have watched the development of these procedures, with the intent of testing which ever procedure(s) looking favorable for forestry purposes.

Both time and the appropriate data sets did not permit testing any of these procedures. A detailed description of the change detection work will be found in the RAP section of this report.

Table 2.7-15 RANKING OF MULTITEMPORAL DATA SET

Best N Channels	Training Classification Accuracy (%)	
	Best 2 of 4 channels per date*	Other 2 of 4 channels per date
14	99.0	99.4
13	98.8	99.4
12	99.0	98.9
11	98.8	98.8
10	98.8	98.8
9	98.4	98.6
8	98.4	98.9
7	98.5	98.1
6	98.0	98.0
5	98.2	97.7
4	98.0	97.1
3	97.7	96.7
2	95.5	94.6
1	70.4	74.5

* The highest average transformed divergence of the Separability processor was used to determine the "best" channel subsets. The seven supervised training classes included: deciduous forest, coniferous forest, cropland, pasture, residential, clear water, and turbid water.

CalComp Capabilities

Computer-aided analysis of MSS data can be a powerful tool for resource managers if results from such analysis are in a user compatible format. Previously, classification results have been available as photographs, either color or black-and-white or as alphanumeric printouts. Although attractive, photographs tend to be costly and difficult to field annotate. Printouts are cumbersome to handle in the field and difficult to use if minor class changes are required. Both photos and printouts require reclassification if class changes or modifications are required.

Clearly, an improved map display was warranted. Maplike quality and ease of annotating or making minor class changes were important factors for consideration. An opportunity to pursue this activity occurred during the early part of the contract year. The resultant product (example Figure 2.7-3) is a line map produced by a CalComp Plotter. A series of subroutines have been written to convert the information on a LARSYS results tape to a format acceptable to the CalComp drum plotter located at Purdue University's Computer Center.

The CalComp will produce maps at different scales and also with different color codings and boundary symbols. However, the most impressive feature, from a user's point of view, is the familiarity of the classification to a map. Field annotations or changes can be easily added. Erroneously classified areas readily changes by drafting the correction provided the change is a minor one or one that reflects human influence rather than spectral class. For example, young regenerated pine will be classified as brushland because spectrally the classes are similar. To change the classification would be difficult, if not impossible, and a costly proposition. Forest managers have information about regenerated areas, so can easily alter the map to reflect the true nature of the land.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Results from this contract year's activities lead to the following conclusions:

1. In order to attain user acceptance of LANDSAT or other remote sensor data, more training and demonstration material highlighting specific applications will have to be developed.

2. Evaluation of classification accuracy is still a difficult task. Users will be slow to accept computer-aided classifications until they can be convinced of the accuracy of the results.

3. The aircraft wavelength bands which simulate the thematic mapper channel configuration yield reasonable classification results for a forest test site in southwestern Colorado.

4. Results for the multitemporal analysis task are not detailed enough to draw extensive conclusions. More work should be pursued in defining the interactions in the data sets. Currently, these analyses are more art than science.

5. No definite conclusions can be made about the effectiveness of change detection activities for forestry purposes.

Recommendations

1. Applications case studies should continue to be developed. These should highlight activities in timber, range or other areas of resource management.

2. Support should be continued in the area of determining classification accuracy from computer assisted analysis. Support should also be available for work in assigning mapping accuracies to computer classified data. Both these areas are important is user acceptance of the technology is expected to increase.

3. Work should be continued with Thematic Mapper simulated data for other major ecosystems in the country. Specific emphasis of these studies should be directed at defining the type of information that can be expected over various forest conditions given the increased spectral rnage and resolution of the Thematic Mapper.

4. An understanding of the interactions between data sets is an important aspect of multi temporal analysis. More work is necessary in describing and understanding these interactions before repeatable analysis techniques can be developed.

5. Change detection techniques can provide valuable information to resource managers. Support for these activities should be continued. Once procedures are developed, they should be tested over well-documented forest sites.

REFERENCES

1. Berkebile, J., Russell, J. and Lube, B. 1976.
Application Simulation of Man-Machine Techniques for Analyzing Remotely Sensed Data.
LARS Information Note 012376, 73 pp.

2. Snedocor, G. W. and Cochran, W. G. 1967.
Statistical Methods. Chapter 11.
Ames: Iowa State University Press, 593 pp.

3. Steele, R. G. and Torrie, J. H. 1960.
Principles and Procedures of Statistics. Chapters 8, 13.
New York: McGraw-Hill, Inc., 481 pp.

4. Anderson, V. L. and McLean, R. A. 1974.
Design of Experiments: A Realistic Approach. Chapter 2.
New York: Marcel Dekker, Inc., 418 pp.

5. Hoffer, R. H. and staff. 1975.
Computer-Aided Analysis of SKYLAB Multispectral Scanner Data in Mountainous Terrain for Land Use, Forestry, Water Resources, and Geologic Applications.

6. Coggeshall and Hoffer. 1973.
Basic Forest Cover Mapping Using Digital Remote Sensor Data and ADP Techniques.
LARS Information Note 030573.
Purdue University Master's Thesis, 731 pp.

2.8 Analysis of Texas Coastal Zone Environments

INTRODUCTION

Inventorying and monitoring land resources is fundamental to the interests of national, state, and local government. In response to Federal legislation designed to give impetus to programs in land resource management at the state level, NASA/JSC conceived the Regional Applications Project (RAP). The primary RAP objective is:

> "To define, design and develop, and demonstrate a Regional Land Resources Inventory System(s), which utilizes in part information extracted from remotely sensed data for the inventory and monitoring of regional land resources in support of resource development and use management."[1]

With the present State of Texas emphasis on the development of a coastal zone management program, an obvious and immediate requirement arises for coastal zone remote sensing applications. For this reason, and for technical reasons associated with the natural diversity of coastal environments and socio-economic dynamics, the Texas Coastal Zone was the first problem area to be addressed in the RAP.

In concert with other efforts to reach the RAP objective, a Supporting Research and Technology (SR&T) program with the Laboratory for Applications of Remote Sensing (LARS) was initiated as part of RAP. Two general technical objectives were proposed in the LARS/SR&T effort to initiate a remote sensing inventorying and monitoring capability in Texas. These were:

1. Determining the feasibility of replicating the Bureau of Economic Geology (BEG) environmental planning units using computer-aided classification of LANDSAT data, and

2. Development of a change detection procedure.

The first objective was addressed during CY75. CY76 was a continuation of the LARS investigation for RAP. The main thrust of the CY76 effort was the implementation and comparison of change detection techniques for identifying and monitoring changes of interest occurring in the Texas Coastal Zone.

The test area for this effort was located in the Matagorda Bay, Texas estuarine system. A detailed description of this coastal zone area is contained in references 2 and 3. Specific test sites were those areas covered by the Austwell; Port Lavaca, E.; Port O'Connor; and Pass Cavallo, SW USGS 7½ minute Topographic Quadrangles.

APPROACH

LANDSAT data were used as the data input for all computer analysis during this effort. The following data were utilized:

Scene ID	Date of Collection	Collection Platform
1127-16260	November 27, 1972	LANDSAT-1
1289-16261	May 8, 1973	LANDSAT-1
1505-16230	December 10, 1973	LANDSAT-1
2034-16200	February 25, 1975	LANDSAT-2

The applicable portions of each of these data sets were overlayed and registered to ground control points to produce a four date temporal data set for each of the 4 quadrangle areas included in the study. These data sets were prepared such as to produce line printer output at a scale of 1:24,000. A non-supervised approach was used to produce a single date classification for each date for each quadrangle area. These classifications were used as base line data. Various change detection techniques were applied to these data sets to test the performance of the individual techniques.

Reference data utilized to support the analysis of the LANDSAT data included 1970, 1971 and 1975 color and color infrared aerial photography, the Geologic Atlases compiled by the Texas Bureau of Economic Geology, and Spectral Environmental Classification overlays produced by Lockheed Electronics Corp./JSC.

PROCEDURES

Preprocessing

The four frames previously described were geometrically corrected and spatially registered to produce an individual four date multitemporal data set for each of the four quadrangle areas.

The registration process was accomplished using the following procedure. First, an area which was covered by the eleven original USGS quadrangles of interest was registered utilizing a biquadratic fit and nearest neighbor interpolation. Second, a coarse geometric correction was applied to this scene. Third, the four original individual quadrangles were precision registered to ground control points.

A more detailed description of the registration process follows. Initially, a base or reference run is chosen. The reference scene identifier* to which all other scenes are registered and the three additional scene identifiers are shown below:

NASA ID	LARSYS ID	DATE
1127-16260	72072100	November 27, 1972
*1289-16261	73126200	May 8, 1973
1505-16230	73126501	December 10, 1973
2034-16200	75000200	February 25, 1975

The automatic numerical integration correlation algorithm of the LARS Image Registration System was utilized to acquire checkpoints. Locations were determined to be the areas of peak correlation in each area (4096 pixels) checked. After the checkpoints between the reference and each of the overlaying scenes were acquired, a least squares regression was conducted to produce a six coefficient biquadratic fit. The resultant fit had an RMS error of less than .500 for the accepted input checkpoints. The checkpoints were selected for land areas only and were well distributed throughout these areas. The three additional scenes were then registered to the reference scene. In the reference scene the area of interest had the following line and column coordinates:

LINE		COLUMN	
First Line	1000	First Column	1
Last Line	2330	Last Column	1634
Line Interval	1	Column Interval	1

Next, a coarse geometric correction was performed on the 16-channel registered data set. Utilizing the latitude of the area to be corrected, this process accomplished three tasks: (a) the image was adjusted in aspect ratio and rotated to true north, (b) the data were deskewed and adjusted for approximately 1:24,000 scale output from the line printer;

nearest neighbor interpolation is utilized in the output pixel determination and (c) a precision registration or "fine correction" was performed.

Checkpoints of land features found in the reference scene image as well as the appropriate USGS quadrangle were identified. A minimum of 20 well distributed checkpoints per quadrangle was required. The quadrangle maps were then individually digitized and the associated measurements translated into lines and columns of the line printer (the line printer is rated at eight lines and ten columns per inch). The checkpoints were then subjected to a least squares regression analysis producing the coefficients (six each per line and column) of the biquadratic fit. The biquadratic fit was utilized to produce the final output image. A 0.500 RMS error was the threshold for checkpoints used in the nearest neighbor interpolation. The upper lefthand corner of the line printer output matches the upper lefthand corner of the quadrangle map. This procedure was completed for the Austwell; Port Lavaca, E.; Pass Cavallo, SW; and Port O'Connor quadrangle areas.

Change Detection

Four techniques for detecting change were selected for evaluation. All four techniques require temporal data which have been spatially registered prior to analysis. Due to a lack of data, the Austwell quadrangle overlay does not include complete coverage for the November 27, 1972 and December 10, 1973 data.

A brief description of the four change detection methods follows:

(a) Post Classification Detection - Remote sensor data are independently classified at both times t_1 and t_2 using whatever analysis techniques are appropriate. A direct change comparator is then used to code changes in a resulting change classification for each pixel. For example, if a two class classification is obtained, e.g., (1) land, and (2) water, then four change comparator results are possible: (1) land to land, (2) land to water, (3) water to water, (4) water to land. An output product is made showing the subset of changes which is of interest to the user.

(b) Delta Data Change Detection - This method is based on differencing (subtracting) remote sensor data from one time and another creating a new difference (called Delta) data set. Multispectral classification techniques are used to analyze and classify change in the Delta data. It is assumed that no-change conditions will result in nominally zero values and these will be combined into one general class. Only one classification is required with this technique rather than two for direct change detection.

(c) Spectral/Temporal Change Classification - The aggregate of all spectral channels from the two dates is used in a change-no change classification in which the classifier is trained to separate out the

different categories. Certain of the spectral/temporal classes would be due to all objects in the scene which did not change and others would be related to those which did change. This technique requires only one classification but it could be an eight-channel classification.

(d) Layered Spectral/Temporal Change Classification - This method is similar to (c) except that the layered classifier is used to perform a hierarchical classification from general classes down to specific classes. For example, change and no-change would be separated in one decision then further classifications would be performed to identify specific change.

Evaluation of the change detection techniques consisted of comparing techniques b, c, and d to the post classification change detection output. This approach was taken since adequate reference data did not exist to allow for a direct visual (photo interpretative) analysis of change which had occurred between the various dates.

Initially, it had been planned to conduct the change detection development effort utilizing the November 1972 and December 1973 data sets. Since these two data sets lack only 13 days of being a year apart, the assumption of little or no seasonal changes nor sun angle effects would have been reasonable. This would have greatly simplified the development of procedures to detect permanent change. However, since neither data set had supporting ground truth, this approach was not feasible. Thus, the approach was modified to include the February 1975 data set, which includes supportive ground truth, in place of the December 1973 data. This introduced the problem of working with data sets containing large amounts of seasonal change and sun angle effects in addition to the permanent change of interest. Progress to date has been limited to detecting change within the Port O'Connor quadrangle area using only two dates of coverage:

Date	Scene ID
November 27, 1972	1127-16260
February 25, 1975	2034-16200

Each of the four change detection techniques is discussed in more detail in the following portions of the report.

Post Classification Change Detection

Post classification change detection is based on the post-classification comparison of independently produced classifications of an area. The comparator, as the name suggests, is simply a processor that "compares" two separate standard LARSYS classifications made of each date and generates a results tape in standard LARSYS format. The comparisons are based on a logical array initialized by the user-specified (change) classes. Thus, the change classes derived from this

method are defined rather than identified and the processing is logical rather than spectral[4].

By properly coding the classification results for times t_1 and t_2, the analyst can produce change maps which show a complete matrix of change. In addition, grouping of classification results allows the analysts to observe any subset of changes which may be of interest. Grouping of the spectral classes is usually necessary when defining change classes as the possible class combinations (MXN) can be a prohibitive number depending on the number of classes (M, N) in each classification.

Initial testing of the post classification change detection technique was directed toward the Port Lavaca, E. quadrangle area using the registered data from November 1972 and February 1975. The technique did indicate change was occurring within the Port Lavaca quadrangle area. However, the area being categorized as change was due primarily to seasonal changes and tidal variations. Comparison of simulated false color images of this area supported this inference. Since this quadrangle appeared to contain little permanent change, future work was directed toward the Port O'Connor area.

Simulated false color images of the Port O'Connor area indicated that major changes of a permanent nature had occurred between November 1972 and February 1975. These changes appeared as ranchland burns, spoil areas, coastal changes, and urban development.

Single date unsupervised classifications of the Port O'Connor area using the registered data from November 1972 and February 1975 were produced for comparison to each other. There are several methods utilized by analysts at LARS for obtaining reference spectral classes. In this analysis false color images, produced from a CRT digital display unit, were used to obtain an overview of the Port O'Connor area and to locate areas of interest. Some detailed information about the surface features of the coastal zone were also obtained from these images.

After the false color images were studied, individual data channels were displayed on the CRT to select training sites of land and water areas. Three training sites were chosen to represent the land areas and an additional three were chosen to represent the water areas. The description of these areas is shown below:

Port O'Connor

LARS Run Number	Site	Lines	Columns
72072105	Land 1	96 to 130 x 1	4 to 24 x 1
	Land 2	36 to 90 x 1	30 to 88 x 1
	Land 3	69 to 81 x 1	136 to 158 x 1

LARS Run Number	Site	Lines	Columns
72072105	Water 1	152 to 183 x 1	11 to 52 x 1
	Water 2	110 to 150 x 1	108 to 143 x 1
	Water 3	19 to 48 x 1	170 to 194 x 1

Two cluster analyses were conducted, one combining the three training sites for the land area and one combining the three training sites for the water areas. Because of storage limitations, each cluster analysis was conducted on 10,000 or fewer data points. The number of classes specified during clustering was 13 and 8 for the land and water areas respectively. These numbers were chosen because they appeared to represent the maximum number of spectrally separable classes that existed within the scene. During clustering, a 99.5% convergence value was used to decrease the amount of computer time required if 100% convergence was selected.

The cluster output included a cluster map showing the location of the spectral classes and a punched output of field description cards for each of the cluster classes. The field description cards were obtained using a "minpoint" option of 3. This requires that 3 spectrally similar points must occur together before they can be described as a field. The number 3 was selected since fewer points introduce a large variance within the classes and since most areas of interest are contained within the 180 meter size of 3 data points.

The field description cards obtained from the cluster analysis were input into a statistics processor to obtain the mean spectral response of each cluster class in all four channels and their covariance matrix. These mean values are used to calculate a ratio $A = \frac{V}{IR}$ where V is the relative intensity of the visible wavelengths [(0.5 to 0.6μm) + (0.6 to 0.7μm)] and IR is the relative intensity of the reflective infrared wavelengths [(0.7 to 0.8μm) + (0.8 to 1.1μm)].

By summing the relative intensity values of all four bands the magnitude of relative spectral responses can be obtained as shown in the following equation:

$$\text{Summed response} = (0.50 \text{ to } 0.60\mu m) + (0.60 \text{ to } 0.70\mu m) + (0.70 \text{ to } 0.80\mu m) + (0.80 \text{ to } 1.10\mu m).$$

By observing the ratio A and the summed response, the analyst tentatively identified the cluster classes. These classes were then maintained, pooled or deleted based upon their separability as determined by their ratios.

Covariance matrix and mean vector statistics of these classes were used to classify the quadrangle area. The classification was output using the PRINTRESULTS processor with a threshold value of 0.5. Often in a classification, there are points which in reality are not

represented by any of the training classes. The classifier necessarily assigns all points to one of the training classes. Thresholding delineates these points and rejects them from the printed results. New training fields were hand-picked from the thresholded areas and their covariance matrix, mean vector statistics and ratios calculated.

The statistics of these new classes were merged with the previously calculated statistics and the quadrangle area reclassified.

The Port O'Connor quadrangle was classified into 25 and 22 spectrally separable classes for the November 1972 and February 1975 dates, respectively.

Prior to conducting the post classification change comparison between these two classifications, it was decided to group the spectral classes into information classes to reduce the number of "changes" observed. This was accomplished with the addition of the reference data which consisted of color infrared (CIR) photography (1971, 1975) and "Spectral Environmental Classification" overlays at a scale of approximately 1:24,000. These environmental classifications furnished by LEC/JSC were prepared using CIR photography and ground observations. The '71 photography was augmented by an overlay prepared by LEC/JSC outlining 34 photointerpreted (training) areas.

The initial step correlated the 18 spectral classes (land areas only) of the computer classifications produced from the February 1975 data with the photointerpreted informational classes contained in the Spectral Environmental Classification. These informational classes were vegetated and non-vegetated classes differentiated into subclasses by ground condition; WET/DRY and surface cover. To accomplish this, the environmental classification was overlaid on the computer classification (Feb '75 data only) and the frequency distributions of the spectral classes versus the informational classes were drawn up (Table 2.8-1). Due to the approximate scale of the overlays and some minor discrepancies in the image registration of the data sets the overlay could not be perfectly aligned with the classification. This implied that the distributions would change depending on how the two were aligned. However, this change was regarded as negligible. As shown in Table 2.8-1 only 24 of the total 55 photointerpreted informational classes were relevant to the Port O'Connor area.

The percent distribution of each spectral class in the informational classes was calculated (Table 2.8-2) from which the relationship of only 4 spectral classes (Feb 1, Feb 12, Feb 13 and Feb A) became evident on a one-to-one correspondence. The classes from the environmental classification were then generalized into broad classes on the basis of their ground condition and surface cover.

Table 2.8-1 Point frequency distribution of February '75 spectral classes versus the spectral environmental classes.

	FEB1	FEB2	FEB3	FEB4	FEB5	FEB6	FEB7	FEB8	FEB9	FEB10	FEB11	FEB12	FEB13	FEB14	FEB15	FEB16	FEB17	FEB18	FEB19	FEB20	FEB21	FEB22	FEB23	FEB24	FEB25
HS																									
HF											2	4													6
HE	8	33	78	53	72	166	229	91	104	87	14	1	57	52	5	35		95							
HEg					1		3			5	2		1	9		2		2							
HEt																									
HEb				9			4	8	8	45	19			3		6		16							
HEi	2	5	7	14	438	65	63	49	78	88	16	15	11	110	173	874	44	223							
Mw	7	70	83	64	7	250	342	100	112	34			65	66			3	97							
Mwg			4	11			12	3		1			8	3				1							
Mwt																									
Mwb							1	1	1	6	7					1		1							
WS																		2							
Wsb																									
WR																									
WRb																									
HG	15	60	144	110	124	213	126	19	30	10	1	78	98	43	54	38	6	156							
HGg		19	13	2		3	8	2	2				5	2				2							
HGt																									
HGb																									
Md	25	103	147	513	15	518	733	92	317	91	14		136	76	4	23		351							
Mdg	10	274	410	460	68	859	575	49	200	82	10		291	41	8	12		267							
Mdb		3	13	10	1	24	39	87	53	386	571		21	24		74		29							
WB																									
WBg																									
WBb																									
WE	5	25	6	30			5	4	11	8	5		5					23							
WEd																									
WEc																									
WEm																									
WEb																									
W	2	14	10	8	294	12	18	12	14	21	10	2	2	34	242	492	354	267							
Wc																									
Wt																									
Ws					71							3		13	37	62		31							
TF																									
TFi	14	11	3	1	175	8	1	9	8	9	1	25		15	217	324	79	127							
TFs																									
BS	3	5	6	3		21	5	4		4		70	2	21	100	19		47							
BSbb																									
BSbf																									
BSwt																									
BSrf																									
MMw																									
TF																									
BS																									
BSdb																									
BSa																		70							
BSb												276			64	1		30							
BSf																									
BSt																									
MMd																									
MMu	146	128	31	2	7	16	4	1					4	4	11			84							
MMr													1												
MMa	4	8		14												1		2							
MMb																									

Table 2.8-2. Percentage frequency distribution of Feb 75 spectral classes versus the spectral environmental classes.

PORT O'CONNOR (Feb. '75)

		Feb A	Feb C	Feb 1	Feb 2	Feb 3	Feb 4	Feb 5	Feb 6	Feb 7	Feb 9	Feb 10	Feb 11	Feb 12	Feb 13	Feb 14	Feb 15	Feb 16	Feb 17
Wet	HE		05	03	04	08	04	06	08	11	17	11	10	02		08	10	01	02
	HEg												01				02		
			05	03	04	08	04	06	08	11	17	11	11	02		08	12	01	02
Dry	HG	01	08	06	08	15	07	10	10	06	04	03	01		16	14	08	06	02
	HGg				03	01										01			
		01	08	06	11	16	07	10	10	06	04	03	01		16	15	08	06	02
Wet	HEi	09	12	01	01	01	01	34	03	03	09	08	10	02	03	02	21	19	45
Wet	MW	01	05	03	09	09	04	01	12	16	19	12	04			09	13		
	MWg						01			01	01					01	01		
		01	05	03	09	09	05	01	12	17	20	12	04			10	14		
	Md		18	10	14	15	34	01	24	34	17	34	10	02		19	15		01
Dry	Mdg		14	04	36	43	44	05	40	27	09	21	09	01		41	08	01	01
	WF		01	02	03	01	02				01	01	01	01		01			
			33	16	53	59	80	06	64	61	27	56	20	04		61	23	01	02
	HEb		01								02	01	05	03			01		
Burned	Mdg												01	01					
	Mdb		02			01	01		01	02	16	06	44	85		03	05		04
			03			01	01		01	02	18	07	50	89		03	06		04
Dry	BS		02	01	01	01			01		01				15		04	11	01
Non-	BSa		04																
Veg.	BSb		02												58			07	
			08	01	01	01			01		01				73		04	18	01
	MMu		04	61	17	03		01	01							01	01	01	
	MMa			02	01		01												
	MMr																		
			04	63	18	03	01	01	01							01	01	01	
	W	73	14	01	02	01	01	23	01	01	02	01	02	01	03		07	26	25
	Ws		02					06							01		03	04	03
	TFi	16	06	06	01			14			02	01	01		05		03	24	16
		89	22	07	03	01	01	43	01	01	04	02	03	01	09		13	54	44

The aggregated percentages of the spectral classes according to these groupings helped identify most of the remaining classes. However, for 6 of the spectral classes the percentage totals did not exceed 50% (the criterion used to assign a spectral class to an informational class). It was observed that these classes mainly occurred in the vegetated and non-vegetated classes that were inundated (covered with water). Addition of their percent distributions regardless of surface cover yielded totals exceeding 50. The classes were then identified as inundated or submerged grass and tidal flats. Spectral class Feb C could not be assigned to any informational group because of its widespread distribution in all the informational classes. It was labeled as a confusion class. The final spectral class definition and groupings are shown in Table 2.8-3.

The reference data for the November 1972 classification consisted of CIR photography obtained in 1971 and a photointerpreted overlay containing 34 training areas.

These interpreted areas were approximately located on both the "Spectral Environmental Classification" and the computer classification. Since the two photointerpretative overlays were prepared using different classification schemes, the environmental classification was used to update the names of some of the interpreted areas, e.g., some of them interpreted as transitional areas were renamed grazed woody/herbaceous. From the frequency of occurrence of the spectral classes within these training areas, the spectral classes were identified and grouped (Table 2.8-4).

This aggregation allowed for a more manageable number of classes in the initial change detection technique development. If the spectral classes were not grouped, a classification containing 22 spectral classes could yield 462 possible change classes. Grouping to 6-8 classes reduces to 30-56 the number of change classes. The spectral classes were grouped into 5 and 6 informational groups (Table 2.8-5) for the November 1972 and February 1975 classifications, respectively.

This exercise allowed the analyst to have a good representation of the February classification with regard to the relationships between spectral and information classes. However, since there was no concurrent ground truth for the 1972 data, only inferences could be made concerning the combination of these spectral and informational classes.

The November 1972 and February 1975 classifications which had been grouped into 5 and 6 groups, respectively, allowed for a possible matrix of 25 change classes. Further grouping of the land classes (woody/herbaceous, urban) and the water classes (water, spoil and submerged) in date 1 reduced the number of possible "change" classes to 11. From Table 2.8-6 it is evident that the first 5 classes are actually "no change" classes (i.e., they remain constant in both dates) and the remaining 6 classes define the changes.

Table 2.8-3 Port O'Connor, February 1975

Class Definitions and Grouping

Group		Classes
1.	Urban	Feb 1 - Mmu - 61%; Mma - 2%
2.	Mixed Woody/ Herb. Dry & Wet	Feb 2 - Md - 14%; Mdg - 36%; WF - 3% Feb 3 - Md - 15%; Mdg - 43%; WF - 1% Feb 4 - Md - 34%; Mdg - 44%; WF - 2% Feb 6 - Md - 24%; Mdg - 40% Feb 7 - Md - 34%; Mdg - 27% Feb 9 - Mw & Mwg - 20% HEg - 17% HEi - 9%; Md, Mdg & WF - 27% Feb 10 - Md - 34%; Mdg - 21%; WF - 1% Feb 14 - Md - 19%; Mdg - 41%; WF - 1% Feb 15 - HE, HEg (wet) - 12%; HEi - 21%; Mw, Mwg - 14%; Md, Mdg - 23%
3.	Burned - Mixed woody/ herb.	Feb 11 - Mdb - 44%; HEb - 5%; Mwb - 1% Feb 12 - Mdb - 85%; HEb - 3%; Mwb - 1%
4.	Spoil Banks	Feb 13 - BSb - 58%; BS - 15%
5.	Submerged grass and tidal flats	Feb 5 - W - 23%; WS - 6%; TFi - 14%; HEi - 34% Feb 16 - W - 26%; WS - 4%; TFi - 24%; HEi - 19% Feb 17 - W - 25%; WS - 3%; TFi - 16%; HEi - 45%
6.	Confusion class	Feb C
7.	Water	Feb 18 Feb 19 Feb 20 Feb A Feb B

Table 2.8-4 Port O'Connor, November 1972

Class Groupings

Group	Classes
1. Urban	1A, 1
2. Grassy areas and chapparal (woody/herb.)	2 3 4 5 7 8 9 10 12 13
3. Spoil	B C
4. Submerged grass and tidal flats	15 16
5. Water	11 14 17 18 19 20 21 A D

Table 2.8-5 Class Groupings

November 72	February 75
1. Urban	Urban
2. Woody/herb.	Woody/herb.
3. Water	Water
4. Submerged	Submerged
5. Spoil	Spoil
6. -----	Burn

Table 2.8-6 Classes used in Direct Change

	November 72	February 75
1.	Urban	Urban
2.	Woody/herb.	Woody/herb.
3.	Water	Water
4.	Submerged	Submerged
5.	Spoil	Spoil
6.	Else (All Land classes)	Urban
7.	Else (All Land classes)	Woody/herb.
8.	Else (All water, submg, spoil)	Submerged
9.	Else (All water, submg, spoil)	Spoil
10.	Else (All Water, submg, spoil)	Water
11.	Else (All land classes)	Burn

The post classification comparator was utilized with the 11 classes in Table 2.8-6. The results are displayed in Figure 2.8-1.

Visual inspection of the classification map and its accompanying spectral environmental overlay in Figure 2.8-1 shows a good correlation between the two. The burned areas (represented by symbol B) and the spoil banks (symbol S) along the intracoastal waterway are well delineated. New urban construction denoted by symbol "U" seems in good agreement with areas marked Mmu (Man-made urban) on the overlay. The '+' symbol representing the change class of land-submerge occurs mainly along the shoreline delineating the tidal flats (TFi on the overlay). The no change classes of woody/herbaceous, water, spoil, urban and submerged are shown by the symbols "., (blank), S, A and I" respectively. The blank areas also represent the confusion class of date 2 (Feb. 1975) which was ignored in the change matrix.

The computer time involved in executing the post classification change comparator was minimal; this effort required approximately 7 seconds CPU. However, the fact that two separate classifications must be performed before the comparator can be used should not be overlooked.

Delta Change Detection

The change detection method referred to as the delta method is based on multispectral difference image classification. Whereas the other three methods studied use classifications of the original multispectral data, the delta method requires that the spectral values from two times for each pixel first be subtracted, producing a "differential image" form of the data. The principle behind this approach is simply that changes from one time to another will produce a non-zero result which can be detected by a variety of methods. The points which do not change would remain nominally at zero. Sun angle and atmospheric differences are expected to cause non-zero deltas for no-change conditions; however, the no change classes would in this case have some positive or negative value which can be biased out leaving the change values equally detectable.

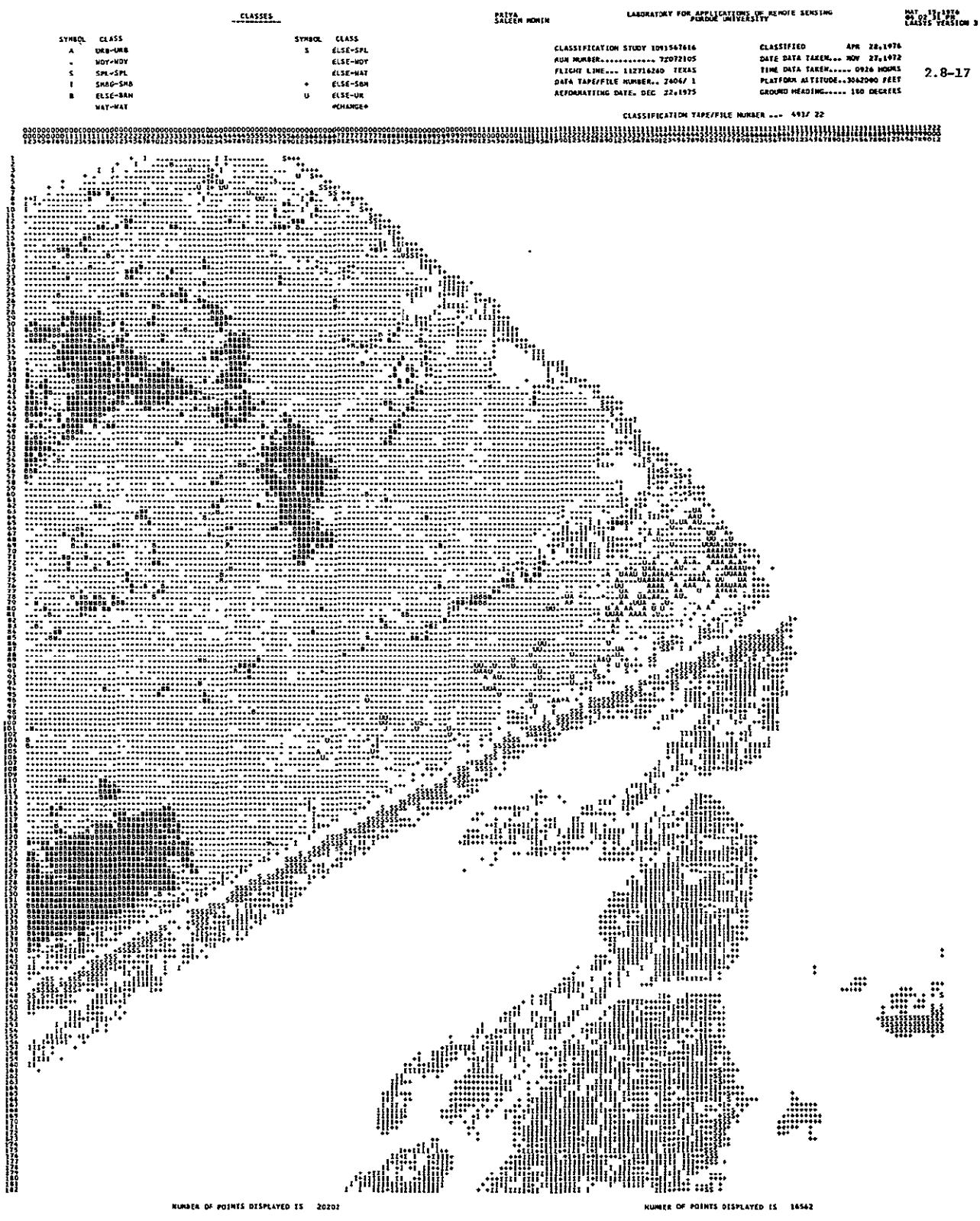
The difference or delta transformation combines two n channel multispectral images obtained at different times and produces a multispectral delta image having n channels. Change can then be detected either by examining the images directly or by first classifying the delta data, then examining the results displayed in image form.

Most image processing systems assume all image samples are non-negative and this is true for the LARSYS system used in this study. To handle negative differences a bias is added to each difference so that the resulting delta image is non-negative. Thus, the complete transformation is simply:

2.8-16

Approximate Scale 1:24,000

LEGEND



2.8-17

Figure 2.8-1. Port O'Connor quadrangle results of the direct change method with the spectral environmental overlay provided by LEC/JSC.

$$\delta X_{ij}^{k} = X_{ij}^{k}\ (t_2) - X_{ij}^{k}\ (t_1) + b_k$$

Where:

X_{ij}^{k} Multispectral value for channel k — i,j = 1,...N Assuming an NxN image

i Row

j Column

k Channel

b_k Bias for Channel k

t_1 First date

t_2 Second date

The method proposed here for detection of change follows a procedure similar to that for classification of multispectral data. The method emphasizes multispectral change classification rather than changes in multispectral classification. Implementation of an automatic scheme in either case requires image registration. A comparison of steps for the two approaches is:

Change Detection by Classification of Delta Images	Observation of Change in Classifications
1. Temporal Image Registration	1. Cluster Analysis at time t_1
2. Delta Transformation	2. Train Classifier for time t_1
3. Cluster Analysis	3. Classify time t_1
4. Train Classifier	4. Cluster Analysis at time t_2
5. Classify Delta Imagery	5. Train Classifier at time t_2
	6. Classify time t_2
	7. Register Classifications
	8. Process Classifications to Determine Changes

In addition to requiring fewer steps the Delta Image method may require less training activity since all features that do not change do not have to be isolated and described. The method proposed here uses non-supervised classification (clustering) to identify samples characterizing various change conditions. These samples are used to train a

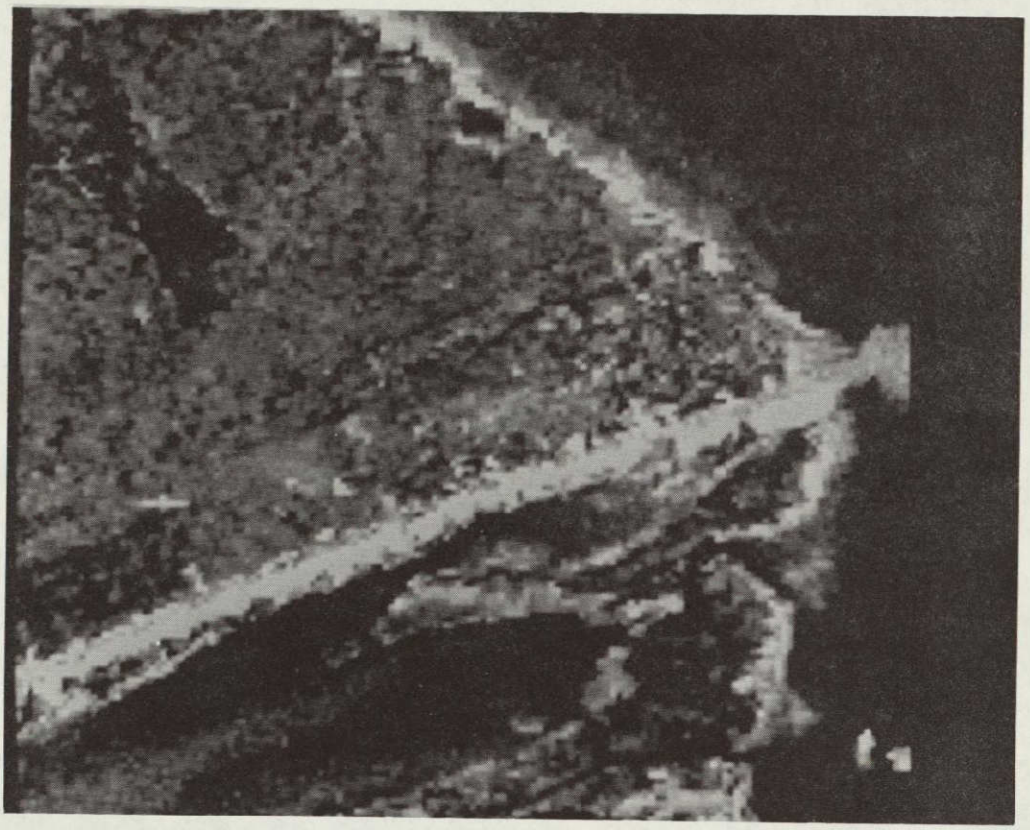
supervised classifier to classify each delta image point over an arbitrarily large area. The results can then be displayed in map form and tabular form the same as for the spectral classification case.

The delta method was tested on the Port O'Connor quadrangle as were the other methods. The registered data from November 27, 1972 and February 25, 1975 was subtracted and a bias of 128 added producing a four channel delta data set. Figure 2.8-2 contains a gray scale image of the delta data. Bright areas represent pixels which have increased in value and dark areas have decreased in value. Medium gray areas are a mix of no-change and small change conditions. The solar elevation angles were very similar for the two dates (34^{o} in November, 38^{o} in February).

The analysis process used was to cluster the delta data to find separable groups of points in the four dimensional delta image data. Several cluster runs were attempted and the thirteen cluster case appeared to give the best separability; thirteen was a reasonable number of change classes. Figure 2,8-3 contains the cluster map created by the LARSYS *CLUSTER program. The mean and variance values for each cluster and the number of points in each cluster are presented in Table 2.8-7 Since the solar elevation angle was essentially the same for both dates it was expected that the no-change classes would be grouped around the data value of 128. This appeared to be the case for in Table 2.8-7 it can be seen that cluster 9 with the largest number of points had mean values of from 122 to 127. Considering only channel 2 clusters 5 thru 9 have means from 131.5 thru 125.4 which is a strong indication that these are no-change clusters.

The next problem was to identify the meaning of each cluster in some quantitative manner. To do this a program was written which compares two classification files and counts the number of occurrences of classes in one file with respect to the other file. Also, an existing linear classifier program was used to extend the cluster results from every other line and column to every line and column for the Port O'Conner quad. The direct change detection results file was then compared to the delta change detection file and counts were computed for each of the twelve direct change classes defined. The results of that count are presented in Table 2.8-8. The table provides a method of quantitatively evaluating the delta classes and enables the clusters to be assigned to specific change or no-change classes.

The majority of points in the quadrangle are no change or no change of interest cases and these were defined first. Clusters six, seven and eight have large counts in the woody to woody and water to water classes and thus were assigned a land no-change class. Clusters nine, ten, eleven and thirteen have large counts in the water-to-water class and were assigned to a water no-change class. Next the change classes were examined and clusters one and two clearly coincided with the else-to-spoil direct change class. Clusters three, four and five were very mixed; however, three had somewhat of a majority of points in else



ORIGINAL IN COLOR

Figure 2.8-2. Gray Scale Image of the Delta Channels Obtained by Differencing February 25, 1975 and November 2, 1972 LANDSAT Data for Port O'Connor Quadrangle. Grassy Woody Areas Burned are Black, New Spoil Along the Waterway is White.

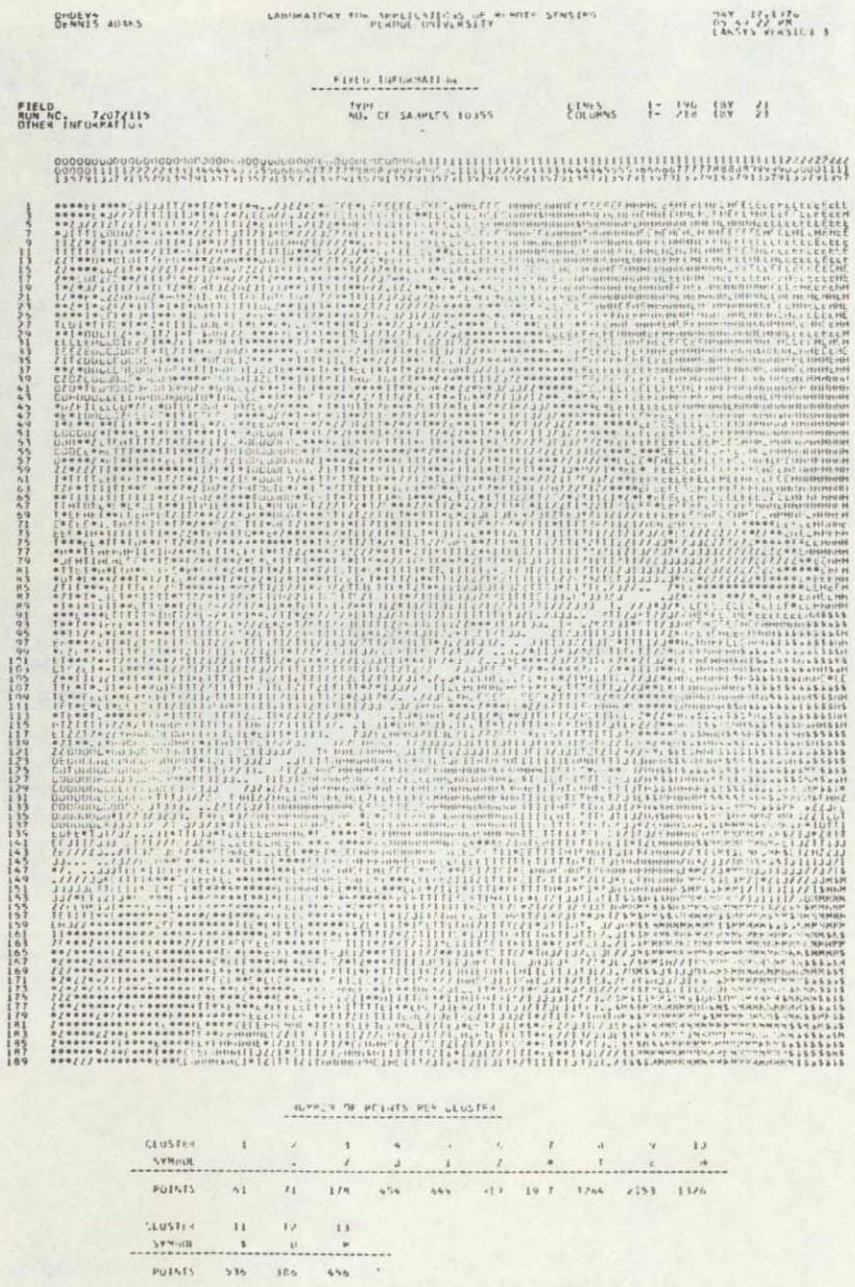


Figure 2.8-3 Delta Cluster Map of the Port O'Connor Quadrangle Using Every Other Line and Sample. Thirteen Clusters were Requested.

Table 2.8-7 Means and Variances for Delta Data Clusters for Port O'Connor Quadrangle.

CLUSTER	POINTS	MEANS CH(1)	CH(2)	CH(3)	CH(4)
1	61	163.80	181.49	176.18	143.02
2	71	146.06	160.42	161.15	139.27
3	178	134.29	144.62	149.72	136.76
4	454	129.83	137.72	139.51	132.29
5	644	125.04	131.51	136.20	131.55
6	919	125.88	131.68	130.53	128.42
7	1907	124.19	129.14	125.99	127.06
8	1264	122.17	127.65	130.00	128.31
9	2153	122.29	125.36	123.81	127.13
10	1326	120.76	121.63	121.80	127.14
11	536	120.28	116.07	116.17	125.83
12	386	119.62	123.49	113.12	119.67
13	456	118.17	111.71	108.72	123.24

CLUSTER VARIANCES

	CH(1)	CH(2)	CH(3)	CH(4)
1	33.47	43.00	37.03	5.43
2	23.57	39.46	23.14	7.90
3	10.46	17.58	16.85	6.60
4	5.29	6.94	10.55	6.47
5	4.13	4.20	5.57	3.83
6	3.75	4.26	2.92	1.98
7	2.91	2.21	2.29	1.51
8	2.32	3.17	3.05	2.04
9	2.51	1.87	2.55	1.57
10	2.80	3.06	2.21	1.17
11	4.50	3.56	7.12	1.54
12	3.36	4.78	10.04	4.37
13	12.51	12.79	12.48	3.15

Table 2.8-8. Counts of the Thirteen Delta Cluster Class Occurrences in the Twelve Direct Change Classes.

	NS- 1/	NS- 2/	NS- 3/	NS- 4/	NS- 5/	NS- 6/	NS- 7/	NS- 8/	NS- 9/	NS-10/	NS-11/	NS-12/	NS-13/
URB-URB I	0	0	0	3	78	46	1	8	0	0	0	0	0
WDY-WDY I	0	0	4	59	604	2460	2853	2288	1494	295	6	86	0
SPL-SPL I	49	67	18	1	0	0	0	0	0	0	0	0	0
SMBG-SMB I	0	15	164	258	226	387	134	638	294	205	8	1	0
ELSE-BRN I	0	0	0	0	0	10	29	36	138	123	5	1191	0
WAT-WAT I	0	0	9	380	87	529	2822	305	4581	4833	1685	72	1325
ELSE-SPL I	177	100	38	3	22	4	0	3	0	0	0	0	0
ELSE-WDY I	0	0	0	0	21	40	15	97	26	17	3	2	0
ELSE-WAT I	0	0	0	59	31	68	40	26	21	16	19	16	94
ELSE-SBM I	0	42	337	391	426	388	111	212	89	83	22	97	5
ELSE-UR I	0	0	2	42	47	14	0	2	0	0	0	0	0
CHANGE I	7	76	195	245	349	426	171	700	220	118	20	54	5

to submerged so was assigned to that class. Cluster four was badly mixed but contained half of the else-to-urban points so was assigned that class. Cluster five was mixed including woody-to-woody, else-to-submerged and else-to-urban. This was assigned a mixed category. Finally, cluster twelve was clearly related to the else-to-burn change and was assigned thus. In this manner an empirical change class assignment was made for the delta cluster classification. Each class was assigned a symbol and the delta classification file was printed out. Figure 2.8-4 contains this printout; blank was used for all water to water classes.

Visual inspection of Figure 2.8-4 shows that a good general agreement exists between the post classification change result and the delta result in that no-change water land interfaces are relatively similar and the burn areas are in very close agreement. Pixels classified into the mixture and urban change classes are very widely dispersed throughout the land and marsh area of the quadrangle. This may be related to the inherent difficulty of reliably discriminating between certain urban and non urban classes with only the four bands of LANDSAT 2, i.e., it may be the result of a limitation of LANDSAT rather than the change detection method. A quantitative evaluation of performance was computed from Table 2.8-8 using the class groupings discussed above. Table 2.8-9 contains the grouped results. The agreement is seen to be in the 50 to 60% range except that the no change classes do somewhat better.

Although the quantitative agreement with post classification change detection is only fair the delta method may have application as a rapid, simple and low cost method of preliminary change detection. The image of the delta channels in Figure 2.8-2 reveals all the change that is in the scene and proper interpretation could possibly allow the desired change information to be extracted. The burn and spoil changes are clearly seen as very dark or bright areas. Construction changes can be picked out since they are usually single bright pixels. Confusion with bright pixels due to spoil bank expansion or exposure of submerged bright sand can usually be differentiated by knowledge of the context of the scene. A color rendition of the delta image can be provided which should enhance the information content of the image.

Timing measurements were made for delta classification of the Port O'Connor quadrangle. The quad contains 36,764 pixels. Computations were done on an IBM 360 Model 67 system.

	Time (sec. total CPU)
Delta Transformation	11
Clustering (13 clusters, every other line and column)	2957
Linear Classification (every line and column)	112
	3080

2-8-24a

Approximate Scale 1:24 000

LEGEND

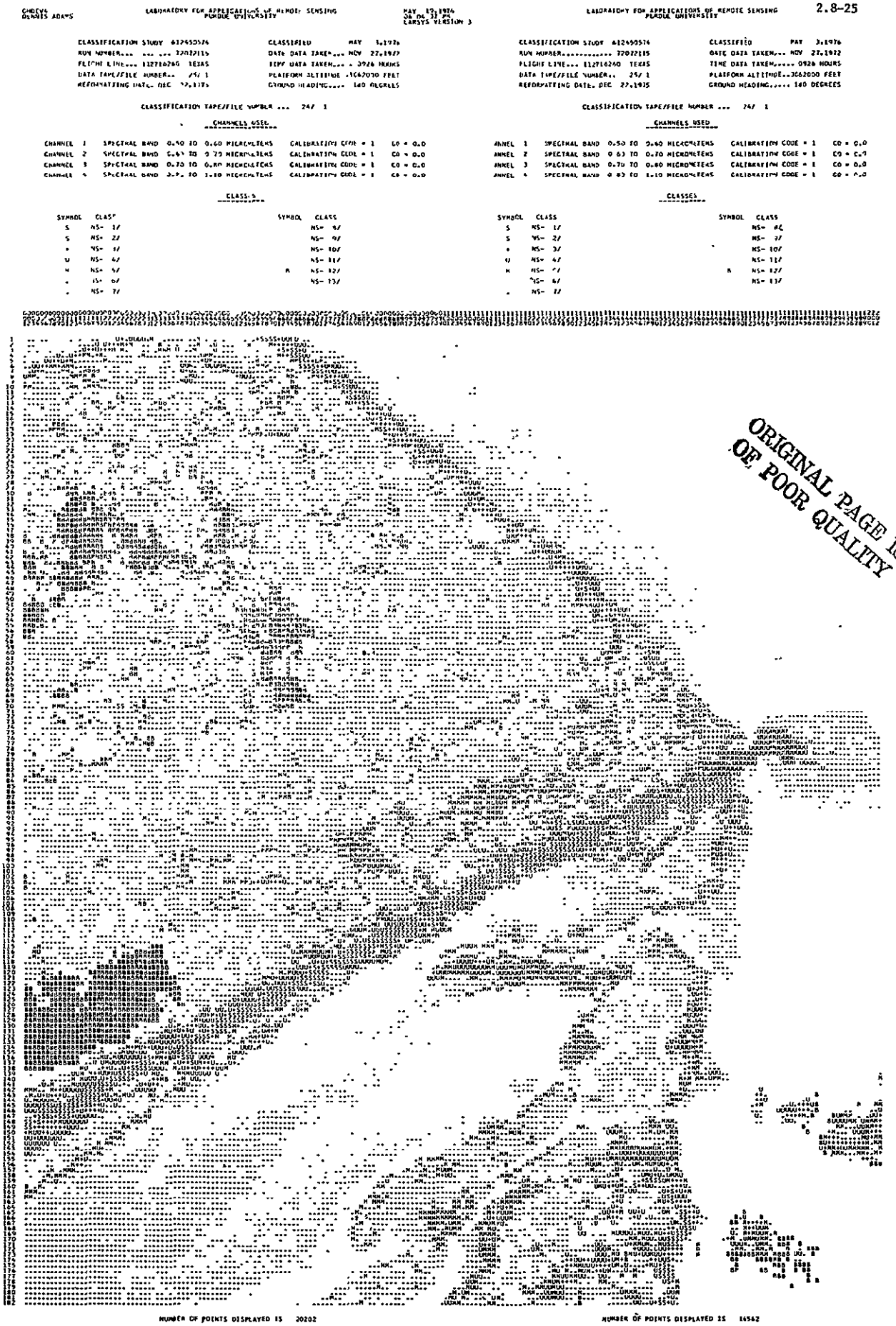


Figure 2.8-4. Delta change detection classification for Port O'Connor quadrangle. Spectral environmental overlay provided by LEC/JSC.

Table 2.8-9. Performance of Delta Change Detection Method Referenced to Direct Change Results.

Direct Change Class	Delta Change Class						
	Else-Spoil	Else-Submerged	Else-Urban	Mixed	Land No-Change	Water No-Change	Else to Burn
Urban-Urban	0	0	3	78	55	0	0
Woody-Woody	0	4	59	604	7593	1795	86
Spoil-Spoil	116	18	1	0	0	0	0
Submerged-Submerged	15	164	258	226	1159	507	1
Else-Burn	0	0	0	0	75	266	1191
Water-Water	0	9	380	87	3656	12424	72
Else-Spoil	277	38	3	22	7	3	0
Else-Woody	0	0	0	21	152	43	2
Else-Water	0	0	59	31	134	150	16
Else-Submerged	42	337	391	426	711	199	97
Else-Urban	0	2	43	47	16	0	0
All other Change	83	195	245	349	1297	363	54
Total	533	767	1442	1891	14855	15750	1519
Percent Class Correct	52%	44%	3%	——	51%	82%	78%

Total Points 36755
Overall Correct 23079
Pct. Overall Correct 63%
Average Class Correct 51.7%

Other minor costs are involved for results display printout but the figure above should be an approximate cost for delta on a per quad basis once the number of clusters is given. The cost is clearly mostly for clustering and if several cluster runs are needed the cost would increase by around 3000 sec. per run.

Spectral/Temporal Change Classification

This change detection technique uses an aggregation of all spectral channels from two dates to produce statistics from an eight channel cluster analysis to identify change. The approach used in this effort made the following assumptions: 1) the areas clustered contained areas that are representative of sites where permanent change had occurred and 2) permanent change can be separated from seasonal change utilizing statistics developed from the spectral data.

In the Port O'Connor quadrangle, five areas were selected for clustering. These areas were chosen as representative samples of the major land use areas. Two areas were delineated in the agriculture regions, one for urban, one along the intercoastal waterway and one on the barrier island. Each area contained 900 points (30 lines by 30 columns). Smaller areas were chosen for clustering so that the number of classes derived from the cluster processor would be manageable. The number of spectral classes requested during clustering is important in determining change classes. As a "rule of thumb," 2n (where n = the total number of expected cover types in both dates) was used in specifying the number of cluster classes requested. Since approximately 5 cover types were expected for each date, a minimum of 20 cluster classes was established. If the resulting cluster classes were found to have large variances (i.e., variance in three or more channels greater than 3.00), the areas were reclustered using 25 classes. Punched statistics containing the means and covariances for each of the cluster classes were obtained as output.

Initial clustering indicated that the potential number of training classes possible from 5 sites would exceed the limitations of the software system. Thus, only 3 of the 5 candidate areas were clustered in the analysis of the Port O'Connor quadrangle.

The three areas used for clustering were:

1. Lines 11-40, Columns 6-35.

2. Lines 63-92, Columns 94-123.

3. Lines 147-176, Columns 112-141.

Area 1 is located in the northwest portion of the quadrangle and contains mostly agriculture and rangelands. Area 2 is located in the vicinity of the town of Port O'Connor. Area 3 is located in the upper area of the barrier island.

Each of these areas was clustered individually, with 20 cluster classes requested. Eight channels, four from each date, were used and the convergence level for clustering was set at 98 percent.

Potential "change" classes were identified using one of three criteria:

1. The ratio ($A = \frac{V}{IR}$) of the mean values for each of the two dates differed more than 0.30.

2. The magnitude (sum of relative spectral response for each of the four channels) for each of the two dates differed more than 25.0.

3. Both the ratios and magnitudes varied greatly.

Tables 2.8-10 and 2.8-11 list the means and variances and the magnitudes and ratios calculated for area 1, respectively. Six potential "change" classes were identified in area 1. These were cluster class numbers 11, 13, 16, 17, 18 and 20 shown in Table 2.8-11.

A similar procedure was followed for identifying potential "change" classes in areas 2 and 3. For area 3, however, it was necessary to request 25 cluster classes in order to reduce the variances to more acceptable levels. Two classes in area 3 still retained high variances, but a request for more cluster classes resulted in some classes having fewer than 10 points in the class, a poor statistical representation.

The separability processor was used to determine what cluster classes would be maintained. Separability was run on each of the areas individually and the classes maintained, pooled or deleted as required.

For area 1, classes 4 and 5 were combined and class 7 deleted resulting in 18 classes. In area 2, classes 2 and 3; 8, 9 and 10; and 11 and 16 were combined, respectively. Class 14 was deleted, resulting in 15 spectral classes. For area 3, only class 8 was deleted.

The statistics for each of the 3 areas were then merged into a single statistics deck, resulting in a statistics deck containing 57 spectral classes. These statistics were input into the separability processor and the results used to select the final combination of classes for classification and the best subset of 4 channels for classification.

Based on separability results class 25 (pooled classes 8, 9, and 10 from area 2) was deleted since it overlapped 4 other classes. Class 5 (6 from area 1) was combined with class 28 (class 13 from area 2). Class 12 (14 from area 1) was combined with class 30 (class 17 from area 2). Classes 45 and 46 (13 and 14 from area 3) were also combined. The Port O'Connor quadrangle was classified using 52 classes and channels 2, 3,

Table 2.8-10 Means and Variances for Area 1

CLUSTER POINTS MEANS

CLUSTER	POINTS	CH(1)	CH(2)	CH(3)	CH(4)	CH(13)	CH(14)	CH(15)	CH(16)
1	39	23.18	18.26	28.79	16.44	19.05	19.79	32.44	15.92
2	65	23.14	18.22	30.77	16.85	18.40	19.35	29.60	15.40
3	61	21.67	16.20	27.97	15.95	17.74	18.11	30.13	15.36
4	50	22.26	16.36	29.06	16.52	17.62	18.50	27.88	14.60
5	70	23.27	18.16	28.01	15.51	17.96	18.60	27.23	14.21
6	37	22.11	16.89	28.68	16.41	16.97	15.00	26.19	13.57
7	49	21.71	16.22	27.04	14.90	16.94	18.20	25.80	13.20
8	55	21.78	16.00	24.15	12.44	17.91	18.51	28.38	14.45
9	34	23.47	18.21	29.74	16.50	17.56	18.06	24.06	12.24
10	19	22.16	16.95	25.26	13.89	16.00	14.79	24.84	12.58
11	42	19.74	14.45	18.21	8.60	17.52	17.83	27.02	13.83
12	26	21.85	16.27	25.62	13.31	17.00	17.77	21.69	11.54
13	42	22.48	17.67	29.31	16.17	15.57	15.40	20.10	10.76
14	26	21.46	16.00	25.23	14.08	15.27	14.27	20.65	10.88
15	19	20.84	15.05	20.47	9.68	16.05	15.74	17.63	9.63
16	44	20.07	14.50	18.34	8.25	16.59	15.98	21.98	11.07
17	52	22.04	17.00	26.10	14.02	14.15	13.00	15.63	7.88
18	54	23.59	19.19	28.93	16.09	15.22	14.76	15.61	8.28
19	22	19.95	14.36	18.23	8.09	13.18	12.50	13.45	6.59
20	94	22.88	18.10	28.90	16.27	13.77	12.32	10.85	5.22

CLUSTER VARIANCES

	CH(1)	CH(2)	CH(3)	CH(4)	CH(13)	CH(14)	CH(15)	CH(16)
1	1.20	2.14	1.48	0.52	1.16	1.27	2.67	0.81
2	1.15	1.55	1.56	0.51	0.71	0.73	1.09	0.56
3	1.59	0.83	1.23	1.05	1.26	1.84	0.98	0.63
4	1.05	0.93	1.20	0.74	1.14	0.46	0.80	0.53
5	0.81	0.92	1.20	0.95	1.06	0.53	0.90	0.63
6	0.93	1.10	1.61	0.69	1.64	0.33	2.05	1.70
7	0.75	0.97	1.92	0.89	1.02	0.33	0.92	0.50
8	1.58	1.19	1.98	1.58	1.20	1.44	3.17	0.96
9	1.23	2.96	2.63	1.11	0.80	1.33	2.54	1.03
10	1.25	1.72	1.20	1.65	1.89	0.18	0.92	0.92
11	1.42	1.86	7.88	3.52	1.33	1.95	2.27	0.87
12	2.14	1.16	1.61	1.50	1.60	0.34	2.86	1.06
13	1.23	1.79	1.88	0.83	2.25	2.88	2.14	0.77
14	1.54	1.20	2.18	2.63	1.96	0.60	1.84	0.91
15	0.70	0.83	2.82	2.12	2.72	3.09	1.80	0.47
16	1.23	1.51	5.25	2.80	1.50	2.67	2.07	0.62
17	1.45	2.08	2.36	2.02	1.90	1.22	2.39	1.52
18	1.42	1.59	2.18	1.03	2.48	2.68	1.26	1.15
19	1.57	1.39	6.37	3.71	1.97	2.36	8.55	3.30
20	1.33	2.35	2.80	1.51	1.67	1.85	2.24	1.01

Table 2.8-11 Ratios and Magnitudes for Area 1

	NOVEMBER 1972		FEBRUARY 1975	
CLUSTER	MAGNITUDE	RATIO	MAGNITUDE	RATIO
1)	86.67	.92	87.2	.80
2)	88.98	.87	82.75	.84
3)	81.79	.86	81.3	.79
4)	84.20	.85	78.6	.85
5)	84.95	.95	78.0	.88
6)	84.09	.86	71.73	.80
7)	79.87	.90	74.14	.90
8)	74.37	1.03	79.25	.85
9)	87.92	.90	71.92	.98
10)	78.26	1.00	68.21	.82
11)	61.0	1.28	76.2	.87
12)	77.05	.98	68.0	1.05
13)	85.63	.88	61.8	1.00
14)	76.77	.95	61.07	.94
15)	66.04	1.19	59.05	1.17
16)	61.16	1.30	65.62	.99
17)	79.16	.97	50.66	1.15
18)	87.8	.95	53.87	1.25
19)	60.63	1.30	45.72	1.28
20)	86.15	.91	42.16	1.62

14, 15 (bands 5 and 6 for each November 1972 and February 1975). Results of this classification are shown in Figure 2.8-5. Table 2.8-12 lists the 52 classes and their assigned change classes.

Four areas of changes were identified from the statistics:

1. Vegetation to Soil
2. Vegetation to Burn
3. Soil to Vegetation
4. Water to Land

Table 2.8-13 lists these change classes and the number of points in each class. Further evaluation of the statistics may reveal additional more subtle changes.

The class counting program was applied to the eight-channel, fifty-two class classification and occurrences with respect to each of the twelve post classification change classes were tabulated. The resulting 12 by 52 matrix is not reproduced here; however, Table 2.8-14 shows for the 52 spectral/temporal classes which of the 12 post classification comparison change classes contained the majority of occurrences. It is seen that the vegetation-to-soil and soil-to-vegetation groups in Table 2.8-13 fall into the woody-to-woody class and the water-to-soil group falls in the submerged-submerged, else-submerged and water-to-water classes. Only the burn change group agrees well with the post classification change results which are being taken as the reference for change detection evaluation.

Approximate Scale 1:24 000

LEGEND

Figure 5. Eight channel spectral/temporal change classification results Port O'Connor quadrangle. Spectral environmental overlay provided by LEC/JSC.

Table 2.8-12. Classification Identifiers and Their Respective Change Classes

Class Number	Classification ID	Change Class	Class Number	Classification ID	Change Class
1	1A	NC*	27	9B	NC
2	2A	NC	28	10B	NC
3	3A	NC	29	11B	NC
4	4A	NC	30	1C	NC
5	5A	NC	31	2C	NC
6	6A	NC	32	3C	NC
7	7A	NC	33	4C	NC
8	8A	NC	34	5C	Water-Soil
9	9A	Soil-Vegetation	35	6C	NC
10	10A	NC	36	7C	NC
11	11A	Vegetation-Soil	37	8C	NC
12	12A	NC	38	9C	Water-Soil
13	13A	NC	39	10C	Water-Soil
14	14A	Soil-Vegetation	40	11C	NC
15	15A	Vegetation-Burn	41	12C	NC
16	16A	Vegetation-Burn	42	13C	NC
17	17A	NC	43	14C	NC
18	18A	Vegetation-Burn	44	15C	NC
19	1B	NC	45	16C	NC
20	2B	NC	46	17C	NC
21	3B	Soil-Vegetation	47	18C	NC
22	4B	NC	48	19C	NC
23	5B	NC	49	20C	NC
24	6B	NC	50	21C	Water-Soil
25	7B	NC	51	22C	NC
26	8B	NC	52	23C	NC

*No Change

Table 2.8-13. Number of Points Per Change Class

Change Group	Class Identifier		Number of Points
Vegetation to Soil	11A		314
		TOTAL:	314
Vegetation to Burn	15A		292
	16A		374
	(18A)		655
		TOTAL:	1321
Soil to Vegetation	9A		298
	14A		199
	3B		46
		TOTAL:	543
Water to Soil	5C		227
	9C		339
	10C		131
	21C		410
		TOTAL:	1107

Table 2.8-14. Maximum likelihood assignment of 52 spectral/temporal classes to the 12 post classification change classes.

Post-Classification Class Change	Spectral/Temporal Classes
Urban-Urban	
Woody-Woody	4A, 6A, 3A, 1A, 2A, 7A, 8A, 9A, 10A, 11A, 12A, 14A, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B
Spoil-Spoil	
Submerged-Submerged	11B, 5C, 6C, 8C, 10C, 12C, 14C, 15C, 16C, 17C, 20C
Else-Burn	13A, 10B, 15A, 16A, 17A,
Water-Water	1C, 11C, 18C, 22C, 23C, 19C
Else-Spoil	
Else-Woody	
Else-Water	
Else-Submerged	2C, 3C, 4C, 7C, 9C, 21C
Else-Urban	
CHG	5A, 9B, 13C

Layered Classifier Approach

INTRODUCTION

The layered or decision tree classifier is essentially a maximum likelihood classifier using multi-stage decision logic. It is characterized by the fact that an unknown sample can be classified into a class using one or more decision functions in a successive manner. This classification strategy can be most easily illustrated by a tree diagram (Figure 2.8-6). A tree generally consists of a root node, a number of nonterminal nodes or decision stages (layers) and terminal nodes. The terminal node corresponds to a terminal decision, i.e. the decision-making procedure terminates, the unknown sample being assigned to the class at that node. However, a nonterminal node is an intermediate decision, its immediate descendent nodes representing the possible outcomes of that decision.

To specify a decision tree uniquely two sets of information are necessary; one set tells how the terminal and nonterminal nodes are linked while the other specifies the decision functions of all the nonterminal nodes. The decision tree can be constructed manually or by an optimized logic tree design procedure.

For the purpose of change detection a hybrid of these two procedures was utilized. Decision trees for each of the two dates were obtained automatically and then manually linked together (with some modifications), thus introducing within the tree a logic for detecting the desired changes.

The flowchart in Figure 2.8-7 illustrates the basic procedure for the layered classifier using the optimized logic tree design processor. The *DISTANCE processor computes the interclass separability (transformed divergence OR the transformed Bhattacharya distance) for selected feature subsets using the statistics as input. All 15 feature subsets of the four channels were used in this case.

The distances were then fed into the *DESIGN processor which uses the "guided search and forward pruning" method[5] and a decision tree t_1 (Figure 2.8-8) was obtained for the November 1972 date.

The February 1975 statistics were split into two decks using the *MERGESTATISTICS processor; one deck contained the burn, urban and vegetation classes and the other water, spoil, confusion and inundated classes. Since changes being investigated occurred only within these two groups of classes, this would cut down on unnecessary nodes in the tree.

The same procedure was used to obtain decision trees t_{2a}, t_{2b} for these two decks. The two trees are illustrated in Figure 2.8-9. The classes contained in the terminal nodes were assigned their informational names which had been derived earlier.

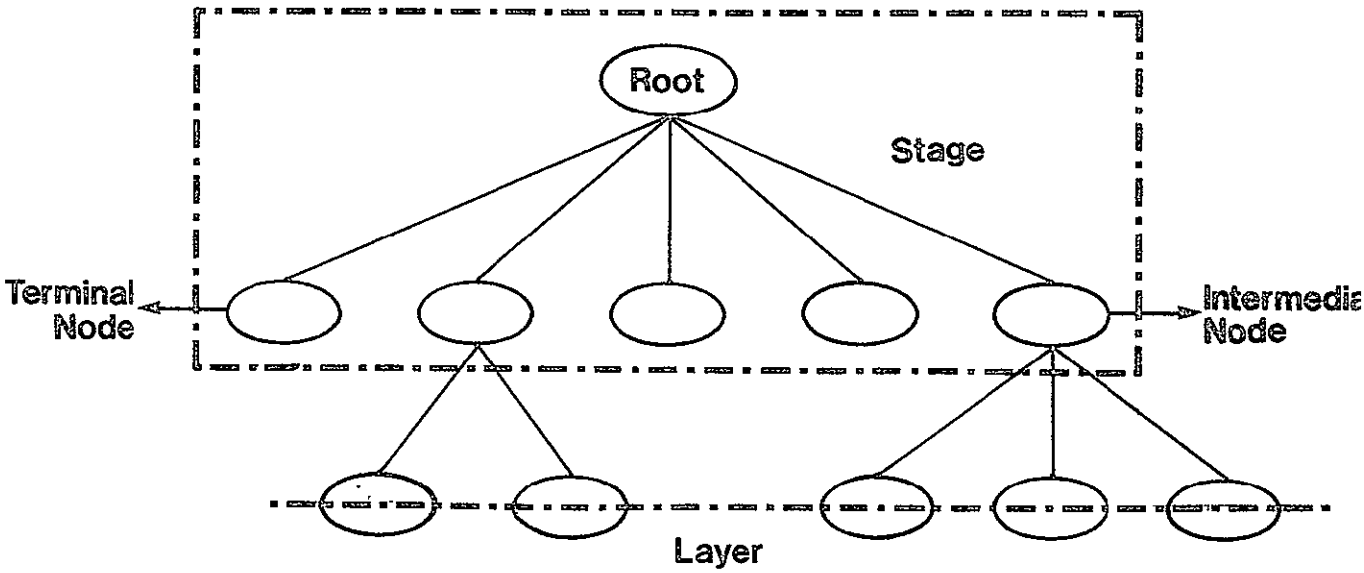


Figure 2.8-6 (above). A simple tree structure. Decision Nodes are drawn as ellipses. The nodes inside the dashed line represent a stage. Nodes along the dot-dashed line represent a layer.[5]

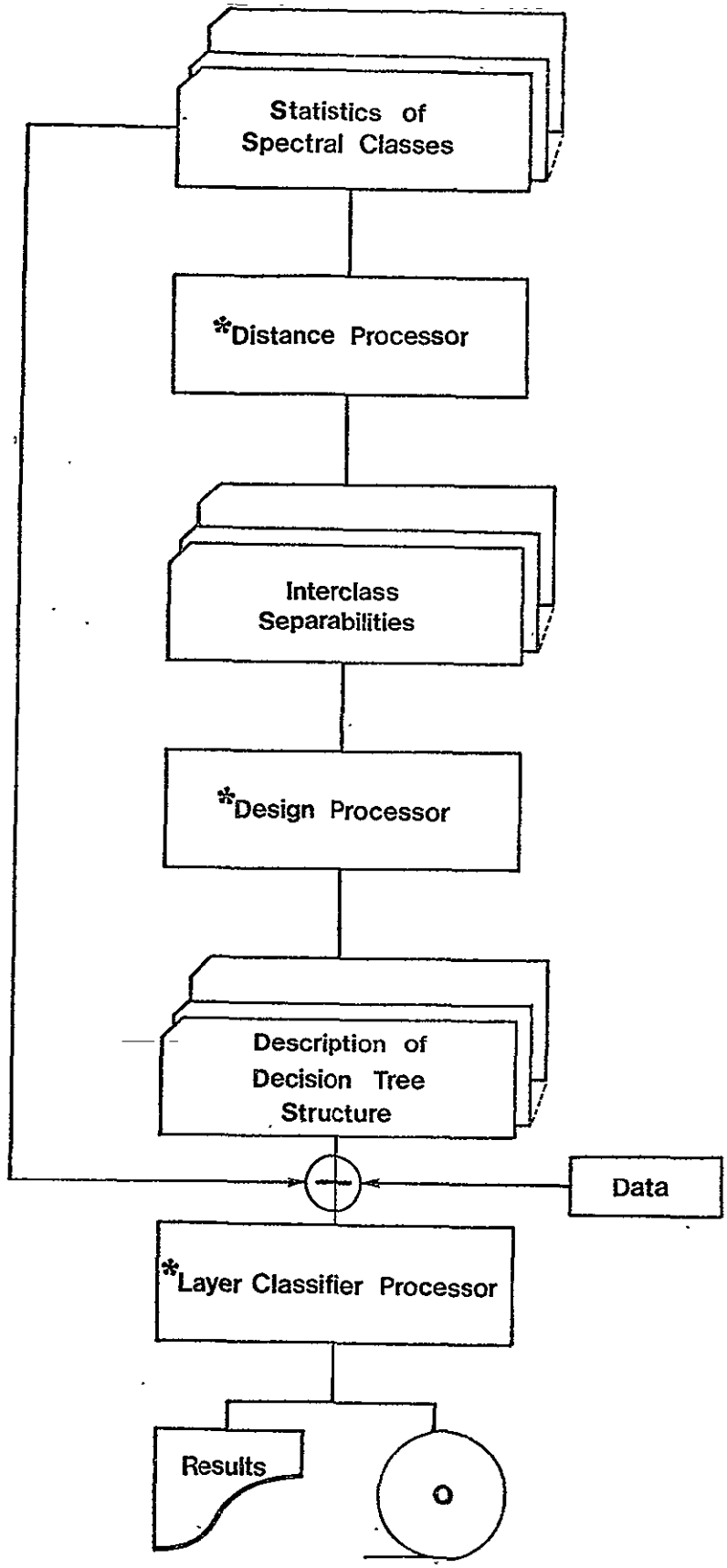


Figure 2.8-7. Input/Output Set up of Decision Tree Procedure with Optimized Logic Approach [Wu]

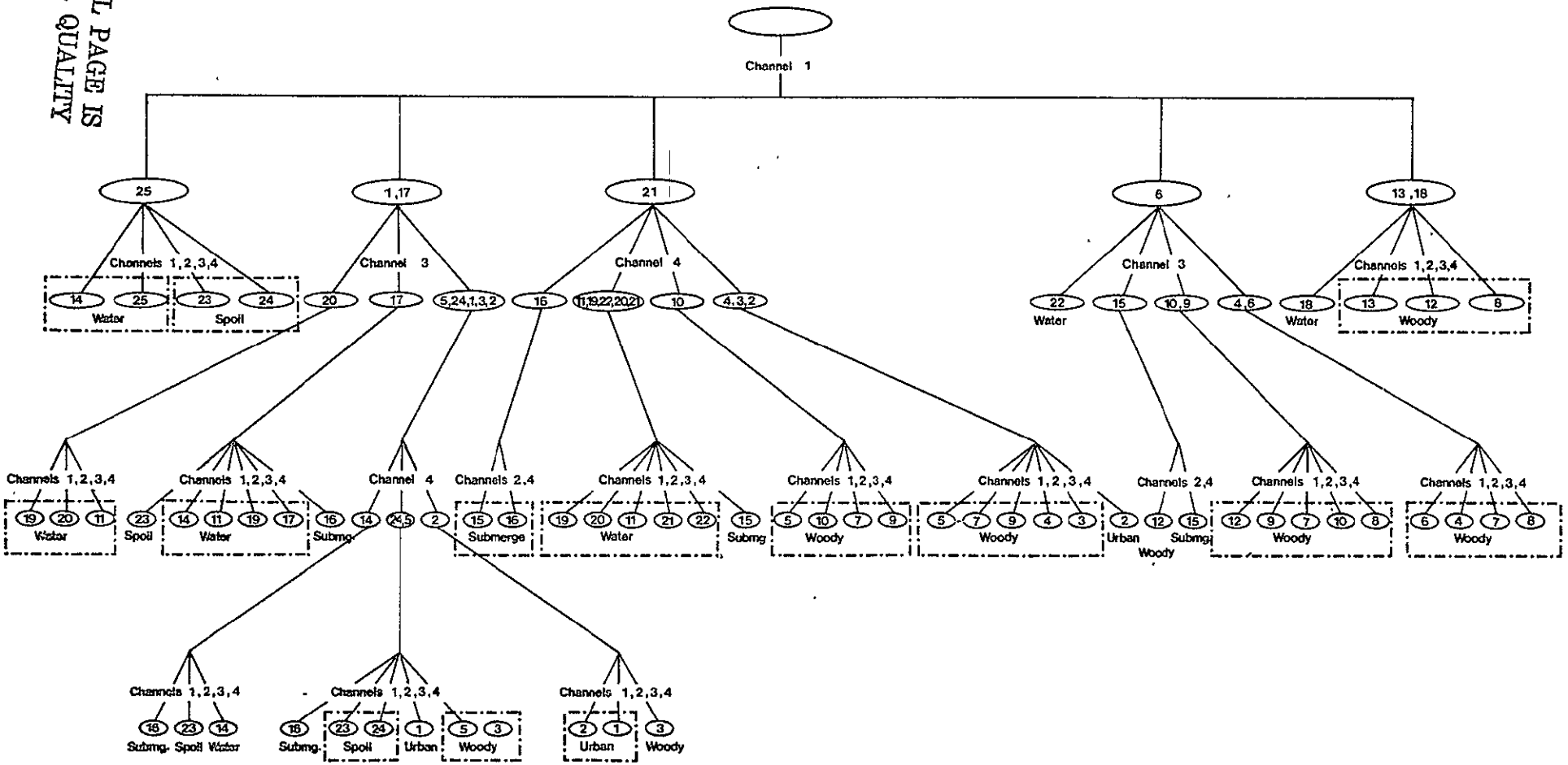


Figure 2.8-8. Decision Tree t_1 Designed from Date 1 Statistics Using Optimized Logic Tree Design. Dotted Lines Indicate Nodes that were Pooled or Deleted in "Pruning" the Tree.

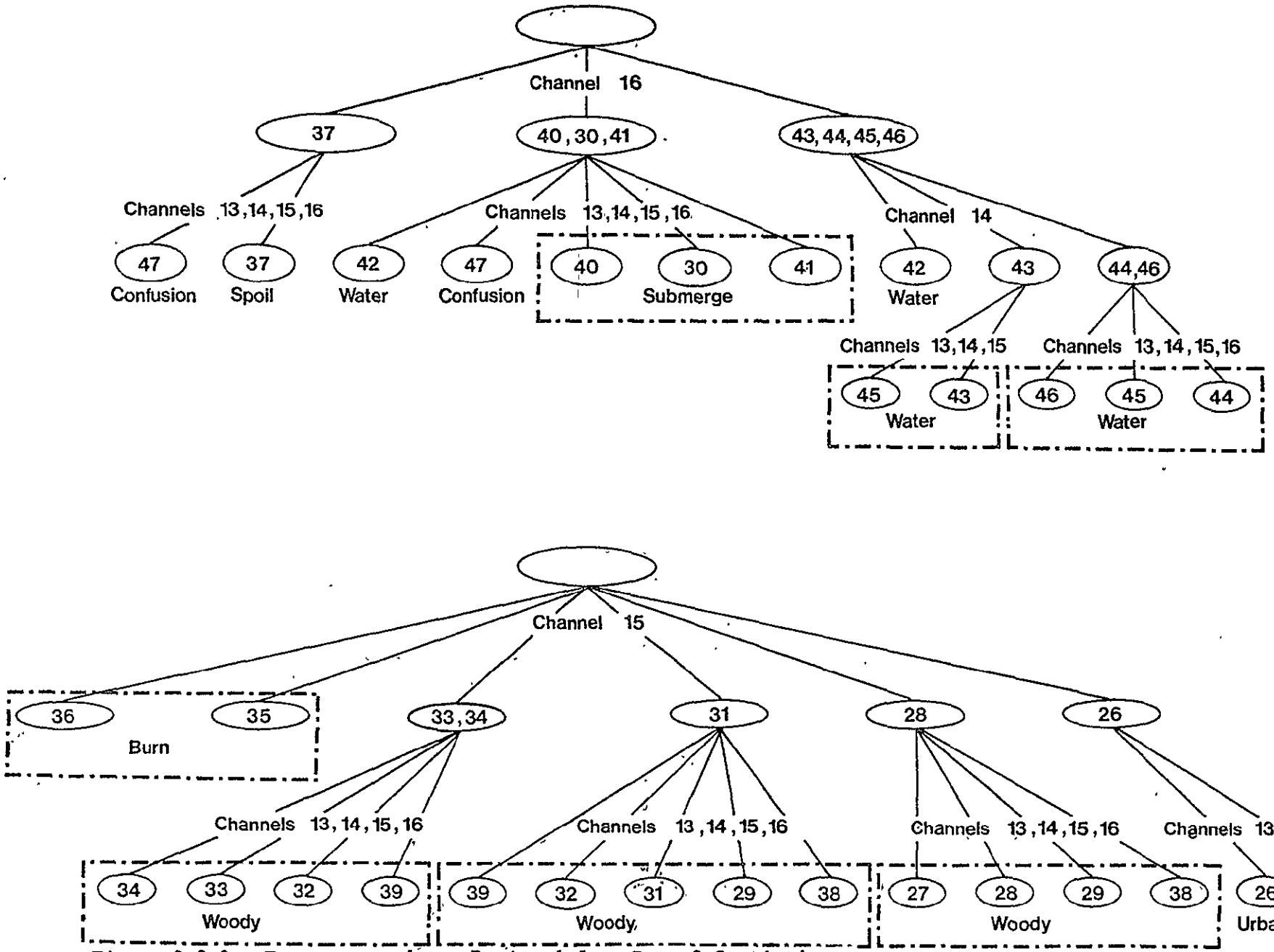


Figure 2.8-9. Trees t_{2a} and t_{2b} Designed from Date 2 Statistics Using Optimized Logic Tree Design. Dotted Lines Indicate Nodes that were Reduced to Singular Nodes or Deleted in "Pruning" the Trees.

Tree t_{2a} was linked to all the terminal nodes of tree t_1 which contained the classes water, spoil or submerged grass/tidal flats. Similarly, tree t_{2b} was connected to terminal nodes of t_1 which contained woody/herbaceous, burn or urban classes. The new tree thus constructed manually was now set up to classify the unknown sample into an informational class (e.g. water) using date 1 (November 1972) statistics and then continue on to see whether that pixel changed to another class (e.g. whether it changed to spoil or submerged grass/tidal flats) or remained the same based on date 2 (February 1975) statistics.

The changes investigated by the tree are listed in Table 2.8-15.

A special fortran program (MOVSTAT)* was used to make the two statistics decks (11/72 and 2/75) compatible for the *MERGESTATISTICS processor, which was then utilized to obtain the eight channel statistics deck required for classification. The experimental *XLAYER classifier was used for classification.

The first classification attempt was unsuccessful as the tree was too large (1221 nodes) to store in the core. In an effort to reduce the size of the tree the various subclasses in the terminal nodes of tree t_1 were grouped at each stage into their major informational class. (These groupings are illustrated in Figure 2.8-8 by the dotted lines.) In some cases if the descending terminal nodes of an intermediate decision node contained only the subclasses of one informational class, all those terminal nodes were deleted. This was justified by the fact that only changes of the major informational classes were being investigated and the decision to classify the unknown sample in that major cover type had already been reached at the intermediate node. After trees t_{2a}, t_{2b} had been trimmed in a similar manner, they were relinked to t_1 to form the new "change decision tree." By this time the number of nodes had been reduced to approximately 700 nodes. The data over Port O'Connor were classified using the 8 channel statistics and the change trees. The classification results are shown in Figure 2.8-10.

---

*Practical difficulties arise in LARSYS when statistics from different dates and locations are to be merged to a physical deck. These difficulties were circumvented by using MOVSTAT which (a) increases the number of spaces for means and covariances to the number necessary for a prespecified number of features; (b) moves the values for means and covariances for date 2 into the correct places in a matrix representing double as many features and (c) adds a set of prespecified constants to the class means.

Table 2.8-15. Changes investigated by Decision Tree.

Class in Date 1	Expected change class (Es) in Date 2
1. Water	Spoil, submerged grass/tidal flats, water
2. Spoil	Water, submerged grass/tidal flats, spoil
3. Submerged grass/tidal flats	Spoil, water, submerged grass/tidal flats
4. Woody/herbaceous	Burn, urban, woody/herb.
5. Urban	Burn, woody/herb., urban

2-8-42a

Approximate Scale 1:24 000

CLASSES

SYMBOL	CLASS	WEIGHT	SYMBOL	CLASS	WEIGHT
U	1	0.021		0	0.021
U	1A	0.021	U	FEB 1	0.021
.	2	0.021	.	FEB 2	0.021
.	3	0.021	.	FEB 3	0.021
.	4	0.021	.	FEB 4	0.021
.	5	0.021	I	FEB 5	0.021
.	7	0.021	.	FEB 6	0.021
.	8	0.021	.	FEB 7	0.021
.	9	0.021	.	FEB 9	0.021
.	10	0.021	.	FEB 10	0.021
	11	0.021	R	FEB 11	0.021
.	12	0.021	R	FEB 12	0.021
.	13	0.021	S	FEB 13	0.021
	14	0.021	.	FEB 14	0.021
I	15	0.021	.	FEB 15	0.021
I	16	0.021	I	FEB 16	0.021
	17	0.021	I	FEB 17	0.021
	18	0.021		FEB 18	0.021
	19	0.021		FEB 19	0.021
	20	0.021		FEB 20	0.021
	21	0.021		FEB A	0.021
	A	0.021		FEB B	0.021
S	B	0.021		FEB C	0.021
S	C	0.021			

DRIVER
SALEEM MOMIN

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAY 27,1976
09:47:43.00
LARSYS VERSION 3

CLASSIFICATION STUDY 814584418
RUN NUMBER..... T2072105
FLIGHT LINE... 112716260 TEXAS
DATA TAPE/FILE NUMBER.. 2608/ 1
REFORMATTING DATE- DEC 22,1975

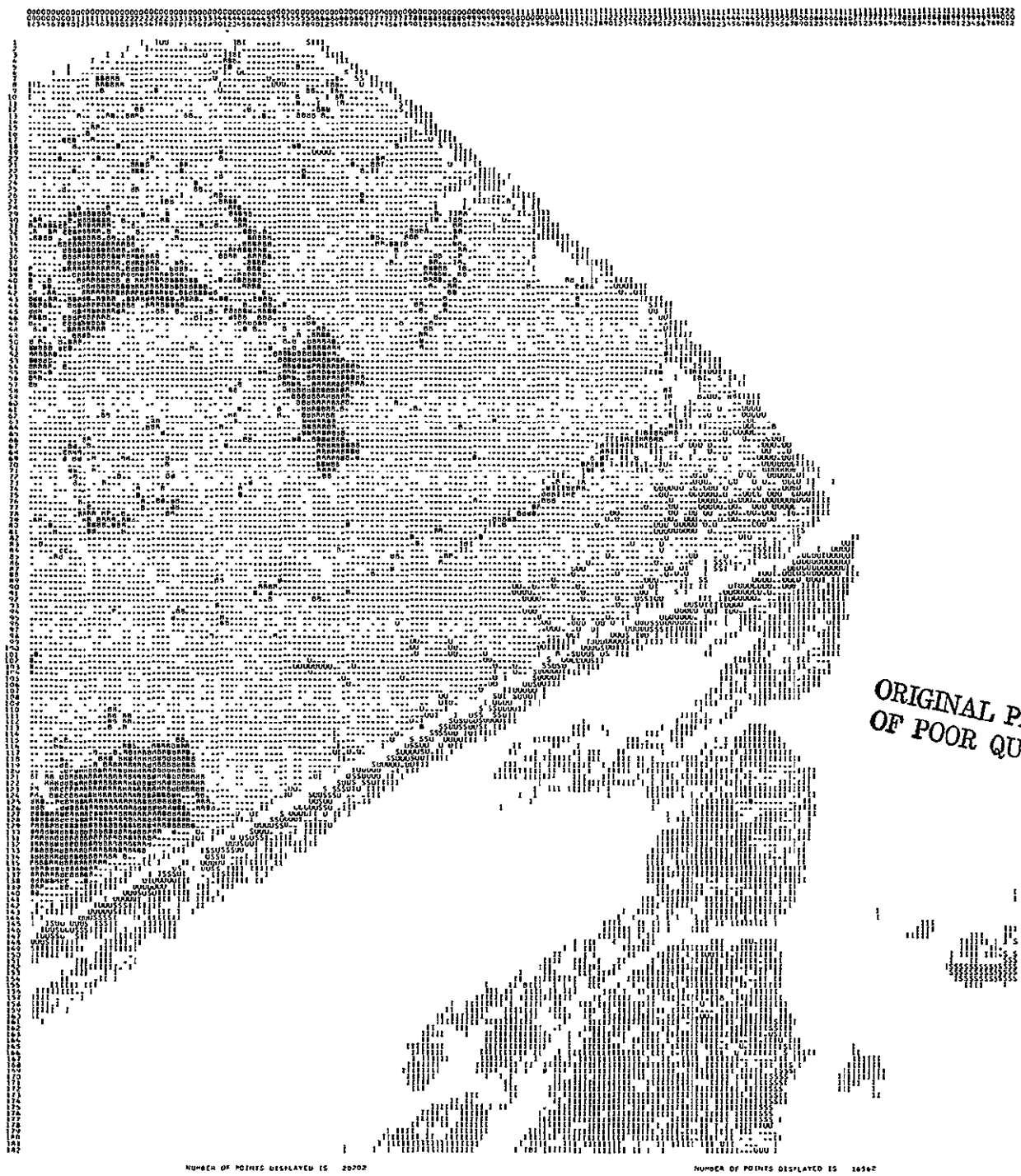
CLASSIFIED MAY 24,1976
DATE DATA TAKEN... NOV 27,1972
TIME DATA TAKEN..... 0920 HOURS
PLATFORM ALTITUDE..3062000 FEET
GROUND HEADING..... 180 DEGREES

2.8-43

CLASSIFICATION TAPE/FILE NUMBER ... 1447/ 1

CHANNELS USED

CHANNEL 1	SPECTRAL BAND	0.50 TO 0.60 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 2	SPECTRAL BAND	0.60 TO 0.70 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 3	SPECTRAL BAND	0.70 TO 0.80 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 4	SPECTRAL BAND	0.80 TO 1.10 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 13	SPECTRAL BAND	0.50 TO 0.60 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 14	SPECTRAL BAND	0.60 TO 0.70 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 15	SPECTRAL BAND	0.70 TO 0.80 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0
CHANNEL 16	SPECTRAL BAND	0.80 TO 1.10 MICROMETERS	CALIBRATION CODE = 1	C0 = 0.0



NUMBER OF POINTS DISPLAYED IS 20702

NUMBER OF POINTS DISPLAYED IS 16562

Figure 10. Layered classifier results, Port O'Connor quadrangle (LARSMAP)
Spectral environmental overlay provided by LEC/JSC.

RESULTS AND DISCUSSION

A qualitative comparison of the "layered" classification results (Figure 2.8-10) with the direct change results (Figure 2.8-1) shows the two maps in good general agreement. The burn areas (B) and the no change land water interfaces are very similar. The urban areas (U) are well identified; but there is some misclassification of the dredged spoil banks along the intracoastal waterway as urban areas due to the similarities in the high spectral response of both the classes. (This problem can be eliminated to some extent by using weights in the classifier and in the tree design processor. However, as yet the weights option has not been programmed into the *XLAYER classifier.) Some small ponds with some submerged vegetation are also misclassified as burn areas because of the spectral similarities of the two classes (both reflect low in the IR).

A quantitative comparison of the Decision tree results and the Post Classification Change results, Table 2.8-16, shows that on a per point basis the two are in agreement, 89% overall, and 77% by class. Some of the misclassification can be attributed to the trimming of the decision trees (which was necessary to accommodate the tree in computer memory). In the case of trees t_{2a} and t_{2b}, when the terminal nodes were deleted, the immediate ascending nodes then became terminal nodes. By definition terminal nodes contain only one class, so all but one of the classes (belonging to the same informational group) were dropped from those nodes. As a result, 5 of the 11 classes of woody/herbaceous and 2 of the 5 classes of water were not used in classification (they were deleted from the tree structure) causing some of the misclassification, especially that of the small ponds, as burn areas.

Table 2.8-17 gives the figures for the CPU time involved in Decision Tree Approach.

Table 2.8-16 Performance of the Layered Classifier Approach as Compared to the Post-Classification Comparison Results.

	Layered Classifier Classes							
Direct Change Classes - Grouped	Woody Urban } Urban	Woody Urban } Woody	Woody Urban } Burn	Water Spoil Submg } Water	Water Spoil Submg } Spoil	Water Spoil Submg } Submg	Water Spoil Submg } Confusion	Total
1. urban-urban + else - urban	229	4	0	0	0	0	10	243
2. woody-woody + else - woody	0	10326	13	0	0	4	27	10370
3. spoil-spoil + else - spoil	245	0	0	0	222	3	12	482
4. water-water + else - water	0	13	4	16855	0	70	76	17018
5. submerged - submerged + else - submerged	89	1520	178	4	0	2740	2	4533
6. else-burn	0	21	1496	0	0	1	14	1532
7. all other change & confusion class	181	1076	83	35	0	101	1110	2586
Percent correct	94%	99.57%	97.65%	99%	46.1%	60%	42.9%	

Total No. of Points 36764

Overall Percent Correct - 89.7%

By Class - 77%

Table 2.8-17. CPU time usage for the decision tree approach.

Processor Used	Time (Sec. Total CPU)
*DISTANCE - Nov. Statistics	29
*DISTANCE - Feb. Statistics (2 decks)	30
*DESIGN - Nov. and Feb.	105
MOVSTAT and *MERGESTATISTICS	150
*XLAYER	192
TOTAL:	506 sec.

SUMMARY AND DISCUSSION

This report discusses the implementation and comparison of four change detection techniques for identifying and monitoring changes of interest occurring in the Texas Coastal Zone. The four methods investigated are: delta change detection, post-classification comparison, spectral/temporal change classification, and layered spectral/temporal change classification.

Since adequate reference data for a thorough evaluation was not available, the post-classification comparison change detection results were used as the standard for evaluating the results from the other three procedures. This choice was based on the assumption before tests were run that this method would provide good results and this assumption was later verified as correct.

In change detection analysis, three kinds of errors can arise: error at time one, error at time two, and error at both times. To precisely evaluate the results of various change detection procedures, adequate ground information must be available to identify any error condition. In the absence of such information, accuracy can be estimated as follows: If P_1 and P_2 are the overall probabilities of correct classification at times t_1 and t_2, if the classifications at different times are independent, and the classes are equally likely, then the change detection accuracy is at least $P_c = P_1P_2$. This expression is a conservative estimate of the accuracy because it counts a no-change situation classified as no-change as correct only when the class is also correctly identified.

Note that for example, if $P_1 = P_2 = .9$ then $P_c = .81$, i.e., at least an 81% correct change classification is achieved if the two spectral classifications are 90% correct. Also note that $P_c \leq P_1$ and P_2; thus the evaluation of a change detection algorithm must be based not only on P_c but P_1 and P_2 as well.

In devising a change detection procedure one property to consider is the relative complexity of the method; all other things being equal, a simpler approach would be more desirable for implementation reasons. On the other hand more complex methods may have more performance potential in the long run. On a scale of increasing complexity the four methods arrange themselves as follows:

1. Delta method. This method requires only a simple subtraction followed by a single classification.
2. Post classification comparison. This method requires two separate classifications followed by a logical comparison.
3. Spectral/Temporal classification. While this method requires only a single classification, it is a vastly more complex one, requiring more classes and probably more features.
4. Layered Spectral/Temporal classification. This method involves not only a complex classification but also a priori knowledge of the logical interrelationship of the classes as well.

During this year, programming for all four of these approaches was implemented and all procedures were executed once on the same data set. The results to this point suggest the following conclusions:

- The delta method may be too simple to adequately deal with all the factors involved in change detection in a natural scene. It may be that too much information is discarded from the data in the subtraction process whereby only the four band difference data remains from the two sets of original four bands. Images created from the delta data may be quite useful, however, in qualitatively assessing change by image interpretation.

- The best results were obtained in this test using the post classification comparison method. This is due in part to the fact that the training procedures required are already routine and quite well understood.

- The latter two methods cannot be ruled out at this point as having great ultimate potential. The layered spectral/temporal change in particular showed best agreement with the post classification comparison results. More complex methods usually require more carefully drawn data inputs together with greater user understanding to achieve their potential.

In summary this effort has provided a good start on the development of practical change detection methods for the Regional Applications Program. However, because of limitations previously pointed out, conclusions must still be considered tentative.

REFERENCES

1. NASA, Lyndon B. Johnson Space Center. 1973.
   Regional Land Resources Inventory and Monitoring Applications System Verification Test Project
   Preliminary Project Plan, Houston, Texas.

2. Bureau of Economic Geology. 1975.
   Environmental Geologic Atlas of the Texas Coastal Zone
   University of Texas, Austin, Texas.

3. Final Report, June 1, 1974 - May 31, 1975.
   NASA Contract NAS9-14016, LARS
   Purdue University.

4. S. G. Luther, M. L. Yanner,
   Machine Processing of Remotely Sensed Data for Change Detection
   LARS, Purdue University, West Lafayette, Indiana.

5. C. L. Wu, D. A. Landgrebe, and P. H. Swain
   The Decision Tree Approach to Classification
   LARS Information Note 090174. Purdue University, West Lafayette, Indiana.

6. H. Hauska and P. H. Swain
   The Decision Tree Classifier: Design and Potential
   Proceedings Purdue Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

7. P. H. Swain, C. L. Wu, D. A. Landgrebe, H. Hauska
   Layered Classification Techniques for Remote Sensing Applications
   LARS Information Note 061275, Purdue University, West Lafayette, Indiana.

2.9 Earth Resource Data Processing Remote Terminal

The Earth Resources Data Processing Remote Temrinal located at the EROS Data Center, Sioux Falls, South Dakota, was in active operation the entire 75-76 fiscal year. Two LARS personnel were assigned to provide the necessary interface with the EDC terminal users. These were Susan Schwingendorf in the area of system support and Barb Davis as analysis techniques specialist. Administrative support was provided by Terry Phillips. By December, 1975, it became apparent that the requested funds would be expended shortly, and a proposal requesting funds for additional computer services was written. As the year closed, further funds were required to allow completion of EDC's projects for the year.

No major problems were reported in operating the terminal. Continued use of the terminal over the nine month contract period was evidenced by the monthly CPU usage, averaging 5.5 CPU hours a month with a high of 15.6 CPU hours in December. Active use was also made of LARS reformatting services, with 20 LANDSAT frames reformatted, 10 areas geometrically corrected, and 4 image registrations performed.

In addition to the above computer facility services, LARS also provided a one week training exercise to two EDC personnel - Michelle Engel and Dale Gehring. Under the direction of Barb Davis, they proceeded through an entire analysis sequence on a LANDSAT data set and learned how to access and use the experimental programs available on LARS' system. These are programs developed in conjunction with various research efforts in the laboratory which have not yet been made part of the LARSYS system.

The EDC terminal is expected to remain in operation through the first half of FY77. EDC has indicated they will be discontinuing use of the LARS terminal as they increase use of their in-house systems.


SEP 1976
RECEIVED
NASA STI FACILITY
INPUT BRANCH